Statistical Estimation and Characterisation Techniques for Use during Accident Response (SECTAR)

N A Higgins, F Burge, T W Charnock and P Teale

ABSTRACT

The SECTAR project investigated the potential of statistical techniques to improve estimates of the extent of radioactive contamination in the early stages of an accident. Two major categories of techniques, geostatistics and Bayesian methods, were investigated and applied to measurement data from nuclear accidents at Tomsk, Chernobyl and Windscale. Trials were carried out using the statistical techniques alone and in conjunction with the traditional assessment methods of atmospheric dispersion and foodchain modelling. Results from this process of 'data assimilation', the combined use of measurements and modelling to make estimates, indicate that several of the methods tested have the potential to be useful to the Food Standards Agency. In particular they will help in the quick and reliable estimation of food restriction areas and in determining areas where sampling effort should be concentrated.

EXECUTIVE SUMMARY

The aim of SECTAR may be succinctly summarised as the investigation of statistical and modelling techniques to make optimal use of the measurement data available after an accidental release of radioactivity.

In the event of an accidental release to the environment, the Food Standards Agency (the Agency) will issue restrictions to prevent the sale of food with radionuclide concentrations exceeding the maximum permitted activity concentration levels in food, specified by the Council of the European Community (CEC 1989a; CEC 1989b; CEC 1990) [*]. The Agency will also impose restrictions on animal feed when contamination exceeds the relevant CFILs. Decisions on the areas where such action is necessary must be taken as soon as possible after the accident, often when there are few results available from monitoring programmes. This carries the risk that the restricted area may be too conservative and the resulting wastage of food may have economic and legal implications. Simple atmospheric and foodchain models of the type used in emergency exercises are not designed to cope with the complications of weather and terrain likely to be encountered in a real release. Conversely, complex models, which may give better predictions, require too much meteorological and other data to be of use at an early stage. The Agency and NRPB[†] identified a need for the development of techniques that can work with the limited data available to make assessments that are more realistic.

The techniques investigated under SECTAR fall into two broad categories: geostatistics and Bayesian methods. Both have the potential to fill the knowledge gap in the early stages of an accident when the emergency plan is no longer adequate but little monitoring data are available. Geostatistics allows interpolative estimates to be made based on the spatial correlation properties of a phenomenon, which are established from a small quantity of data. Bayesian methods have been used to progressively improve or replace process model[‡] predictions using measurement data. These, or closely related statistical techniques, allow the uncertainty on estimates to be assessed, for example geostatistical simulations are used to find the probability that the contamination in a particular area will exceed a given limit.

Use of real accident data has been of great importance to the work of SECTAR, since it not only provides the input data for the techniques but also the criterion against which their success is judged. Data were obtained for the accidents at Chernobyl (Bryansk), Tomsk and Windscale; this allowed the performance of the techniques to be evaluated for a variety of release conditions.

---

[*] These will be referred to as Council Food Intervention Levels (CFILs).
[†] The National Radiological Protection Board (NRPB) became the Radiation Protection Division of the Health Protection Agency (HPA-RPD) on the 1 April 2005.
[‡] A mathematical model of the physical process involved in the transfer of radioactivity through the environment.

Results from these investigations have identified a number of techniques that may, with some refinements, prove useful to the Agency in establishing appropriate food restriction areas in the early stages of an accident. The ability to estimate the uncertainty on predictions should reduce the risk of imposing over-cautious restrictions and allow the more effective deployment of monitoring teams.

ACKNOWLEDGEMENTS

CONTENTS

1   INTRODUCTION

The Food Standards Agency (the Agency[*]) has a requirement to establish the need for, and likely extent of, food restriction zones[†] as soon as possible after an accidental release of radioactivity. The zones must cover all locations where the radionuclide concentrations in food exceed the maximum permitted activity concentration levels in food, specified by the Council of the European Community (CEC 1989a; CEC 1989b; CEC 1990)[‡]. The Agency would like to estimate the extent of these zones accurately so that any additional conservatism introduced to accommodate both practical considerations and uncertainty can be controlled. Currently, the Agency supports the estimation of zones by the use of its own field measurement teams, monitoring from other sources, and both atmospheric and foodchain modelling. However, decisions must be substantially formulated within a very short period after the accident. This time scale is sufficiently short to raise doubts about the potential of any monitoring campaigns to provide robust estimates (ie, estimates which are not excessively conservative) within that time scale. Additionally, the uncertainties associated with existing rapid modelling approaches are well known. It is thus unlikely that the models currently used will provide adequate support for decisions on the appropriate level of conservatism to apply under circumstances that deviate substantially from the situations rehearsed in exercises. Practical atmospheric dispersion models that may be used in the limited time and with the information available will not provide an accurate representation of the ground deposition pattern. The Windscale fire of 1957 (Clarke, 1974) provides an example of the difficulties inherent in modelling a real release.

Following an accident it is likely that much more monitoring will occur than is assumed during exercises. However, the time constraints inherent in responding to an accident and carrying out any necessary chemical analyses will still prevent the Agency from rapidly assessing the extent of the CFIL boundaries using the traditional modelling and measurement approach without the precaution of excessive conservatism.

Monitoring in the early stages of an event is carried out by the site operators to establish the path of the plume, and to identify locations where the emergency countermeasures of sheltering, evacuation and iodine prophylaxis may be appropriate. Such monitoring will therefore be concentrated near populated areas. More generally, the need for rapid assessment will concentrate early measurements in areas easily accessible from a road. Use of results obtained in this manner are appropriate when assessing doses to people from inhalation and deposition, as people may be expected to live or work within a short distance of a road. However, animal and crop production will generally take place at some distance from population centres and the roads between them. Thus, the monitoring used to assess the immediate protection needs of the population from inhalation and external doses is unlikely to provide optimal support for the more

---

[*]Formerly the Ministry of Agriculture, Fisheries and Food (MAFF) had this responsibility.
[†]This is the area over which a Food and Environmental Protection Act (FEPA) 1985 restriction would apply.
[‡] These will be referred to as Council Food Intervention Levels (CFILs).

homogeneous, and therefore more time-consuming, coverage required by the Agency. Aerial monitoring is a valuable technique that will partly overcome this difficulty but it is expensive and will not eradicate all problems. Aerial monitoring only allows detection of a limited range of radionuclides[*] and the limit of detection is higher than for ground based measurements (Bourgeois *et al*, 1995). This may be compounded by a large and potentially varying footprint (the ground area within the detector's field of view) over which the information is averaged. If the size of this footprint is larger than that required by the Agency the area might need to be re-sampled using alternative techniques to reveal the true level of variability in results sampled at the chosen scale (see Section 4).

Thus, even a significant increase in monitoring effort will only have a limited effect on the problem faced by the Agency. It will provide some additional data but not necessarily in the time or form required, by those making decisions on the extent of food restrictions. The conversion of monitoring data into a suitable form for these purposes may be far from straightforward.

It was in the light of the known problems that the Agency embarked on this programme of work with NRPB (now HPA-RPD) to investigate methods for the effective assimilation of information from measurements and simple models. In this context assimilation is the use of mathematical techniques to make optimal predictions about levels of environmental contamination and, if possible, the uncertainty in those predictions through the combined use of all the available sources of information. The requirements of the Agency coincided with a realisation at NRPB that the techniques of data assimilation, the closer integration of measurements and models, offered the prospect of a more efficient and effective use of the available data. Thus for any given level of monitoring, the use of assimilation techniques would enable decision-makers to have a more thorough and well-founded understanding of the situation on which to base their decisions.

SECTAR therefore seeks to maximise the amount and quality of information available to a decision-maker by making use of data from all available sources such as ground and aerial monitoring. Ideally, other data sources such as radar rainfall measurements, process modelling and known ground susceptibilities (ie the importance of soil types on the uptake and persistence of radionuclides in the immediate environment) would be included in an analysis[†]. The use of process modelling has been considered but information on the other options was not available for the three accidents considered in SECTAR. However, an investigation into the potential use of radar rainfall data in the estimation of initial deposition is described in Higgins and Jones (2003). Information on ground susceptibility has been reported in, for example, Nisbet *et al* (1999).

---

[*]However geostatistical techniques, for example, may be used in conjunction with aerial monitoring to infer more effectively the distribution of radionuclides only detected by less extensive ground-based monitoring results.

[†]The last option would allow risk maps to be produced showing areas where there is a likelihood of larger transfers of deposited radioactivity into crops and animals or a longer environmental half life of deposited material.

1.1   Organisation of material

The work undertaken for the SECTAR project depends on the use of measurements from past accidents. Such data exhibit a complexity, which is not present in, for example, the model-generated data used in emergency exercises. The statistical techniques investigated under SECTAR are intended to cope with this complexity. Thus, only the use of real data can provide an adequate test of the statistical techniques under investigation. Section 2 introduces the data obtained for the project and subsequently used in a series of trials designed to test the applicability and usefulness of statistical methods under a variety of conditions. Section 3 provides an outline of the theoretical approaches employed. These techniques are not widely used in radiation protection and it is important to have a broad understanding of the assumptions inherent in their use. This division between the theory and its practical application allows the discussion of the results to be more succinct and to concentrate on the particular features of the data and modelling assumptions that affected the outcome of trials. The trials applied to each set of accident data are discussed in turn in Sections 4, 5 and 6. For intercomparison, some trials have been made with common features using data from the different accidents. However, the particular nature of the data in each case and the use of techniques applicable to those data are also considered. Section 7 discusses the results, highlighting the usefulness of the techniques examined, and considers what methods can be easily applied and where further development work may be appropriate.

2   COLLECTION OF DATA

Considerable effort was expended in trying to obtain monitoring data from past accidents. The SECTAR project is specifically concerned with improving decision making for real events (as opposed to emergency exercises). Monitoring data were therefore a vital part of the project, since, as discussed in Section 1.1, they provide the only way to test the practical potential of the techniques developed. The data are used as both input and as the measure of a technique's success, that is, the technique must be able to reproduce the observed measurement results within an acceptable margin of error. Although the simple atmospheric dispersion models available to the Agency and NRPB (now HPA-RPD) for use during early accident response have been of considerable use in the project, they cannot be used to provide realistic simulations of the measurements likely from real events. Thus, data form exercises, designed to test administrative aspects of emergency response, may be unsuitable for the testing of techniques intended to work with genuine environmental data. This is because exercise data are often unnaturally 'smooth' or include arbitrary 'hot-spots'. Access to real data is thus an essential prerequisite, if the anticipated improvement over non-adaptive[*] modelling is to be demonstrated.

---

[*]An adaptive model in this context will change its predictions as the result of the influence of measurements and by more than a simple scaling of the predictions.

Three major datasets were obtained for use in SECTAR: measurements from Tomsk in Russia after the accident at the Tomsk-7 chemical plant in 1993, measurements of ^{137}Cs from Bryansk in Russia after the Chernobyl accident, and measurements made after the Windscale fire of 1957. However, a sufficiently detailed source of UK Chernobyl data has remained elusive, preventing work on the use of rainfall measurements as supporting data (Higgins and Jones, 2003).

2.1   Tomsk data

The Tomsk accident occurred on 6th April 1993, when an explosion in a facility for handling uranium solution at the Tomsk-7 plant started a short release of radioactive material. Further details of the accident can be found in Shershakov *et al* (1995). Scientists from the Scientific Production Association (SPA) 'Typhoon' Rosgidomet (Obninsk, Russia) took environmental dose rate measurements within days of the accident. The following September a more extensive survey was undertaken using a helicopter-mounted gamma spectrometer to measure ^{106}Ru, ^{95}Nb and ^{95}Zr. The ^{106}Ru results from this second survey provided the basic material for the studies undertaken in SECTAR. Figure 1 gives an overview of the 812 measurement points in the dataset.
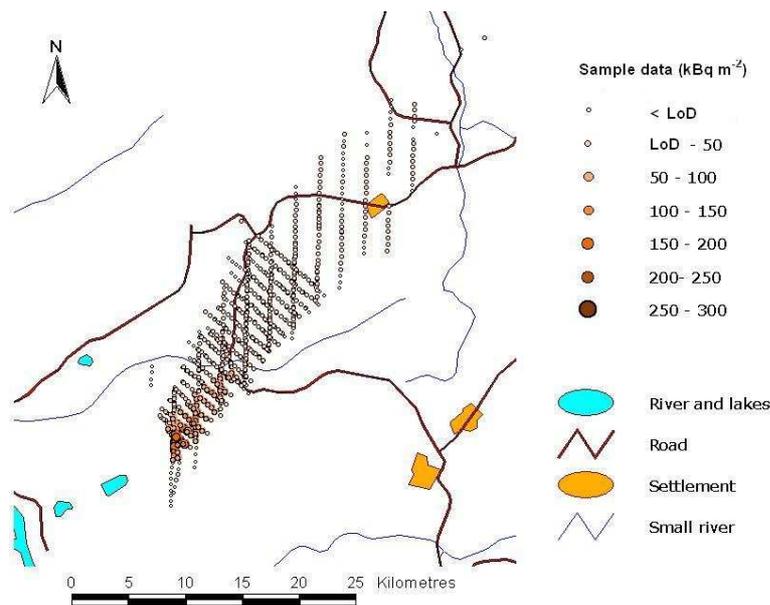
FIGURE 1 Available measurements from the accident at Tomsk including those below the Limit of detection (LoD).

The figure shows that the investigators concentrated on the area of highest deposition closest to the release. Another feature of the data is that, because the helicopter flew at varying heights, the lower limit of detection varies between 3700 and 7400 Bq m^{-2} (100-200 mCi km^{-2}). Some 118 points at the edge of the plume are marked as being below the limit of detection (LoD), however insufficient information was available about which LoD applied at any particular location. Several methods exist for handling such data,

which can add valuable information to what is known as a censored dataset. Such techniques are outside the scope of the SECTAR project but a discussion of their application, to measurements of radioactivity in food, is provided in Daniels and Higgins (2002). It should be noted that food restriction areas based on European Community maximum permitted levels (CEC 1989a; CEC 1989b; CEC 1990) would extend beyond the sampled area. This demonstrates that good estimation at the edge and beyond can be an important consideration when evaluating techniques.

Figure 2 illustrates the output of a simple Gaussian atmospheric dispersion model, and comparison with Figure 1 shows that such a model appears to describe the Tomsk data well. The implications of this for the successful application of geostatistical techniques to this data are discussed in Section 4.
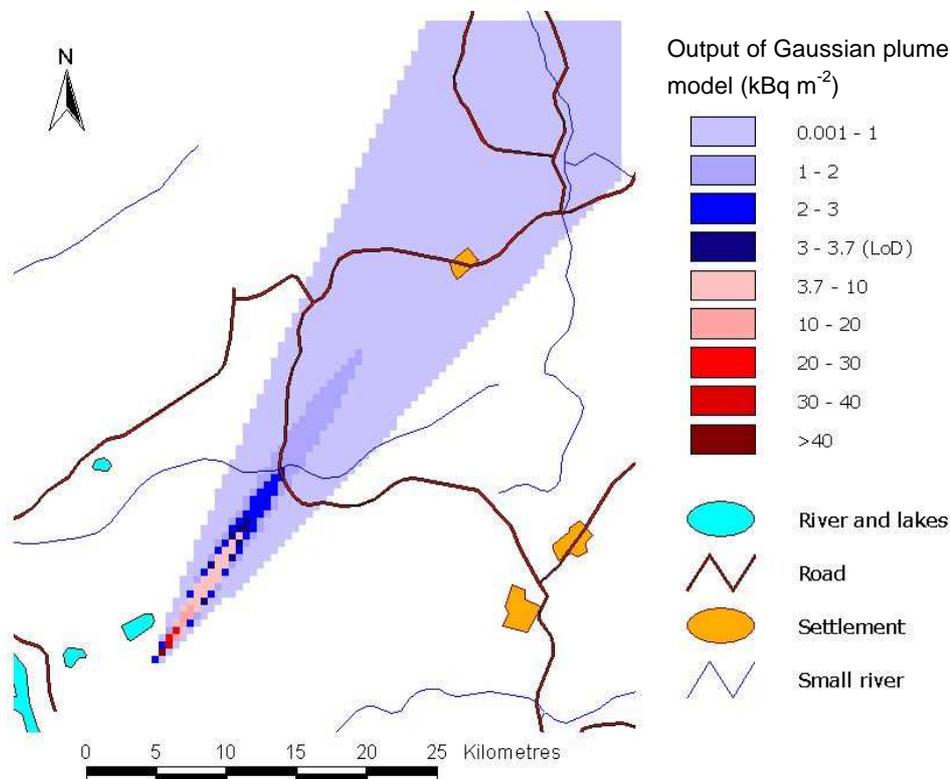
FIGURE 2 Gaussian plume model for Tomsk, using default model parameters and showing predicted values that are above and below the Limit of Detection (LoD) of the measurements shown in Figure 1.

2.2   Bryansk data

The Bryansk dataset, consisting of 1226 measurements of ^{137}Cs in soil, is illustrated in Figure 3. The measurements, supplied by Yatsalo (1997), were made in the Novozybkov district of Russia following the Chernobyl accident. Each measurement represents averaged ground survey results taken in 1992-93. The averaging procedure varies depending on the land classification but for arable land a sample representing an area of 250 x 200m^2 is the averaged result from 20-25 bores, each of which was 20cm

deep. The Bryansk data cover a greater area but have a lower percentage variance than the Tomsk data. Possible causes of this observation are the greater distance of Bryansk from the source (the sampled area is approximately 175 km from Chernobyl), and the longer period that the plume was over the area. The relationship between the variance of a set of measurements and the area represented by each sample value is an important topic in geostatistics and is discussed further in Section 3.1.2.4.
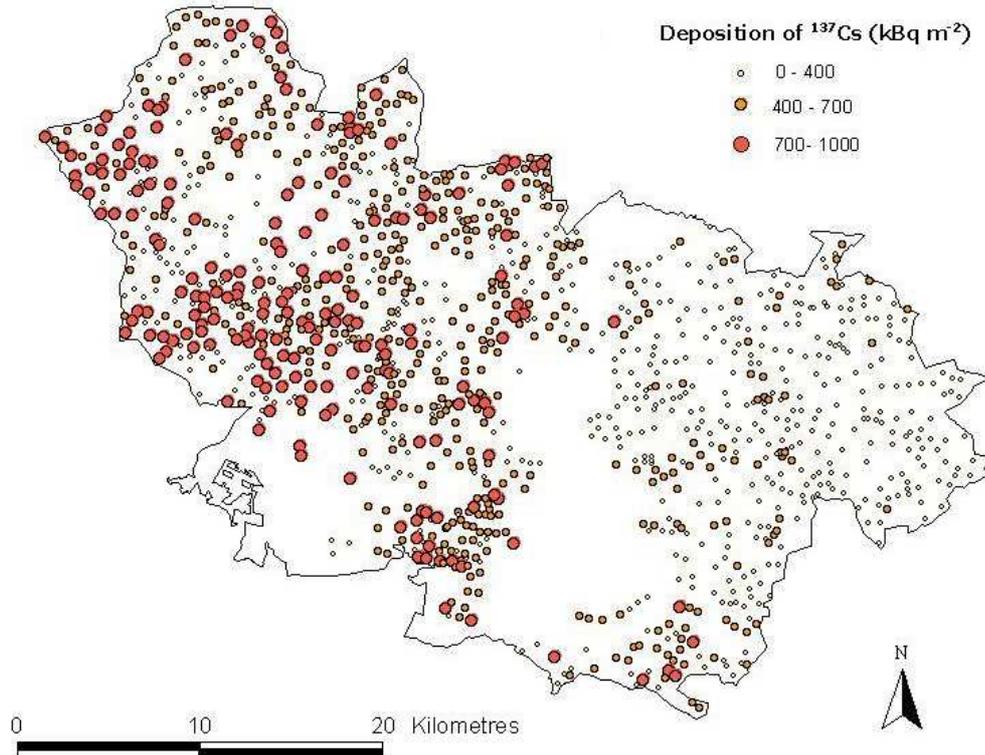
FIGURE 3 Bryansk data set.

2.3   The Windscale database

The Windscale dataset is the largest and most diverse set of post-accident measurements currently available to HPA-RPD. In the weeks following the fire at Windscale Pile No.1 in October 1957, an extensive programme of district surveys (Dunster *et al*, 1958) was carried out to measure radioactive contamination of the environment. These surveys covered the local area in detail and, to a lesser extent, more distant locations. In total, approximately 8000 measurements of a range of radionuclides in 12 major categories were taken. These categories included animal and vegetable products, human thyroid measurements and measurements of the non-living environment. This information has been entered into a database and testing undertaken to determine the accuracy of the transfer from the original card and paper records. The database has also been linked to a Geographic Information system (GIS) to simplify the

selection and use of data and importantly to provide a check on the validity of the reported measurement locations[*]. Details of the database and testing undertaken are provided in Appendix D.
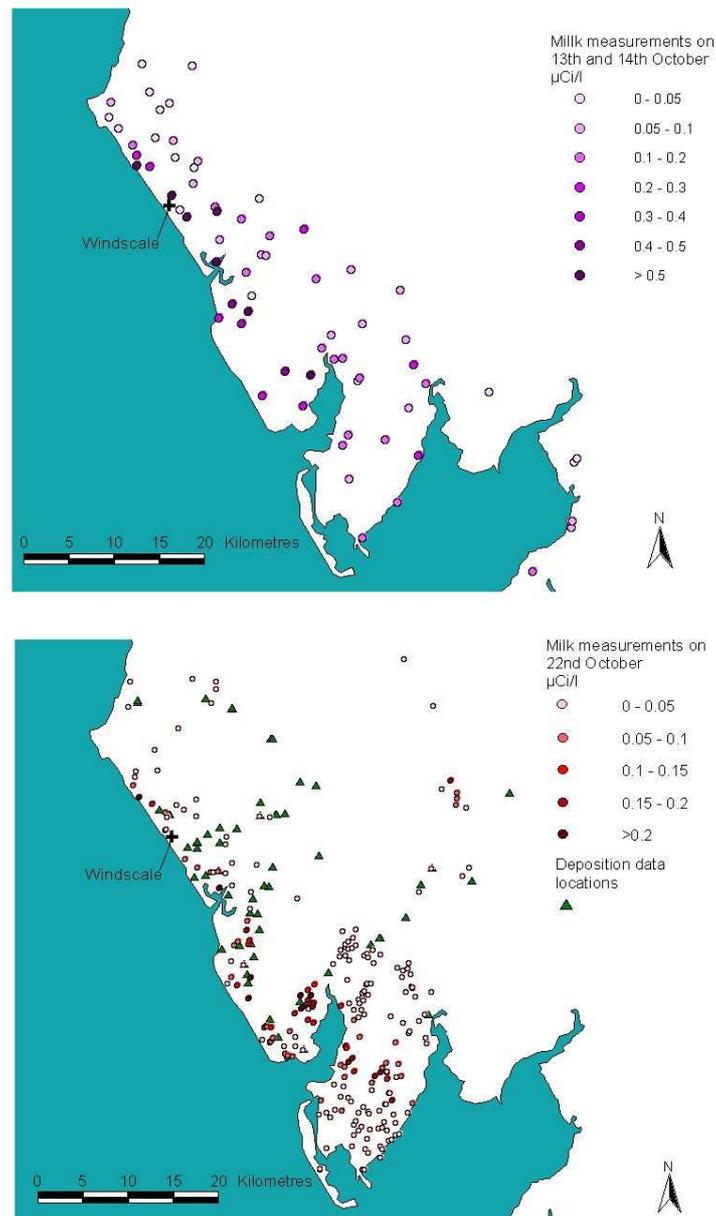
FIGURE 4 Two datasets selected from the Windscale database for use in SECTAR geostatistics experiments (see Section 6)[†].

---

[*]The locations written on the typed records occasionally required to be deciphered due to smudging over time.
[†]At the time of the Windscale accident, activity was measured in Ci, and the milk ban criterion was also expressed in these units. Windscale studies concentrate on predicting the milk ban and measurements are therefore shown in the original units or in multiples of the milk ban activity level. A measurement of 1µCi/l corresponds to 37 kBq/l.

The Windscale data used in SECTAR were selected from measurements of ^{131}I in milk and on grass, with both sample types being thought reliable and reasonably numerous. These data are illustrated in Figure 4. The database also contains many measurements of gamma dose rates and air samples, but further supplementary information is required to make these suitable for use in calculations. For example, the energy response characteristics of the type of meter used, for the majority of gamma dose-rate measurements, is, as discussed in Section 6.5, already available. However, difficulties remain over the timing of the measurements with respect to the release and estimates of the mix of radionuclides released. These need to be resolved before the dose rate data can be used. HPA-RPD is continuing to search for further supplementary information, which may be held by other organisations.

The database has greatly enhanced the usefulness of the Windscale data in the SECTAR project by allowing the rapid selection of data subsets appropriate to particular investigations. The connection of the database to a GIS has also simplified the selection of suitable data based on their spatial distribution, and the display and analysis of the calculated results.

2.4 Summary

Although it was only possible to obtain three data sets that were suitable for use within SECTAR, each has its own very distinctive characteristics. Factors such as release duration, type of sampling (aerial or ground-based), location of sampling relative to source, meteorological conditions, and terrain, are different for each of the accidents. There is a more detailed discussion of these factors for the individual accidents in Sections 4, 5 and 6 respectively but the essential point for the purposes of SECTAR is that they cover a wide range of possible accident conditions. It has therefore been possible to establish the performance of several of the techniques investigated for more than one type of accident, and to determine the conditions under which they are likely to be of most benefit. The demonstrated ability of the techniques to handle this range of conditions supports the belief that they are likely to be useful in the event of a real accident (see Section 7).

3 THEORETICAL BACKGROUND

The techniques investigated under SECTAR may be separated into two main areas: geostatistics and Bayesian methods. Geostatistics is a branch of spatial statistics concerned with inferring the values a variable may have at different locations. At its simplest, it calculates the best interpolation estimates for a quantity based on its spatial characteristics, which are determined from a limited number of measured values. Simple measures of the reliability of the estimates derived by this approach are automatically produced with the estimates. The use of geostatistics also enables more comprehensive estimates of uncertainty in predicted values (and related probability estimates) to be calculated using stochastic simulation (see Section 3.1.4). The

Bayesian approach employed uses measurements to calibrate a process model[*]. The particular implementation of this technique within SECTAR also handles poor performance of the model predictions through the introduction of a 'model inadequacy' term. If the predictions of the best fitting model do not agree well with the measured values, the importance of the inadequacy term increases to compensate. Thus, in the extreme case, the model predictions are replaced with an estimate based on the data alone (effectively a simple geostatistical (kriging) estimate, see Section 3.2). The different methods are used with a common aim of trying to extract as much information as possible from a small amount of data, and ideally providing an estimate of the uncertainty in the predictions generated. They both therefore have the potential to be useful to the Agency in making decisions based on the few measurements available in the early stages of off-site accident monitoring.

3.1 Geostatistics

Geostatistics developed from the work carried out by the mining engineer D.G. Krige during the early 1950s in the field of ore reserve estimation (Krige, 1951). These ideas were formalised and extended by G. Matheron, who was the first to suggest the name 'geostatistics' (Matheron, 1962). As the theoretical development of geostatistics has continued, it has found practical application in many areas. In addition to the mining industry and earth sciences, the techniques have been used in a wide range of environmental disciplines where there is a correlation in space and/or time between measured values.

The techniques of geostatistics apply to measured values that have an observed or assumed correlation due to their spatial proximity to one another. At their simplest, the techniques do not use any information other than the measured values. This can therefore be considered as an approach that is entirely complementary to the use of a simple empirical process model that takes no or little account of the measured values which the model is attempting to represent. However, the techniques of geostatistics also offer ways of combining other information, including information from process models, with direct measurements of the quantity of interest.

The techniques of geostatistics can be used to quantify the high level of spatial uncertainty inherent in predictions based on a small number of sample values. Geostatistics relies on the theory of 'regionalised variables' to represent, using a random function, the large-scale spatial correlation and small-scale irregularity observed between sample values (Chiles and Delfiner, 1999). In this theory, the regionalised variable is the quantity of interest, and it has a spatial (and possibly a temporal) dependence. The values measured at each location are considered particular realisations of a random function representing the regionalised variable. A random

---

[*] The term "process model" is a shorthand term for a model that represents the processes involved in the movement of radioactivity between different parts of the environment, as distinct from a statistical model of the relationship between measured and possibly modelled values. The distinction can be quite minor in practice as environmental models often have (statistically fitted) parameters that are only defined by the fitting procedure and the context of their use in a particular model.

function representation is chosen because the phenomenon is simply too complicated to be modelled with a deterministic function. A probabilistic approach is therefore the only practical option. Thus, although there is only one sample environment, it is assumed to be drawn from a population of independent but identically distributed sample environments which each observe the same spatial correlation between individual measurements. Thus, if the processes involved in generating the regionalised variable could be re-run under identical conditions, it is assumed that the values measured would be different from those found in the previous trial.

In terms of the work of SECTAR, the quantity of interest may be, for example, the deposition, d, of a particular radionuclide. This has a spatial dependence (and also a temporal dependence which is neglected here for clarity), and deposition is therefore a regionalised variable d($\mathbf{x}$) which has a value at all points, $\mathbf{x}$, over the region of interest. The measured deposition values are samples of the regionalised variable at particular locations. The physical mechanism by which deposition occurs is dependent on many factors – meteorology, the form in which the radionuclide was released, terrain and so on, and clearly cannot be specified as a deterministic function unless simplifying assumptions are made. Further progress thus necessitates the modelling of the regionalised variable d($\mathbf{x}$) as a random variable; that is, a particular selection from a random function D($\mathbf{x}$).

An important consideration in geostatistics is whether the phenomenon under study can be modelled using a *stationary* random function. Strict stationarity requires that the probability distribution $Z(x)$ for the variable of interest is translationally invariant in space ie on an appropriately large scale the mean and all higher moments of the distribution, calculated within an appropriate neighbourhood of a point, are the same for all locations. The condition which is more usually applied is 2nd order, or weak, stationarity. This only requires the first 2 moments of the distribution, the mean and covariance, to be translationally invariant – the mean *m* is constant and the covariance C(*h*) depends only on the separation of points, *h*, not on their absolute locations *x*, as shown in equations 1 and 2 respectively:

$$E\{Z(x)\} = m \qquad\qquad (1)$$

$$E\{(Z(x)-m)(Z(x+h)-m)\} = C(h) \qquad\qquad (2)$$

In reality, it is unlikely that the data can be well represented on a *global* scale (ie over the area covered by the entire dataset) by a stationary random function. For example, the amount deposited at a location by a dispersing plume decreases with distance from the source as the plume broadens and deposited material is lost from the plume. As indicated above, the mean surface concentration produced by a plume dispersing overhead will depend on factors such as weather conditions and terrain, and will vary from area to area and with the size of the area considered.

Methods for dealing with this global non-stationarity will depend on the scale over which the phenomenon varies, relative to the distances between measurement locations. If the area over which variation takes place is large, it is possible to define *local* search

areas over which the mean can be considered constant. Estimations may then be carried out using the data in each area separately (ordinary kriging, Section 3.1.2.2). However, when the change in the mean is apparent in areas which are small in comparison to any search area containing enough data points for estimation, it is necessary to adopt methods which deal with the change or trend explicitly (kriging with a trend, Section 3.1.2.3).

In addition to the problem of stationarity, directionality is likely to be important. Thus, the spatial correlation of a set of measurements is often found to vary with direction, that is, it is anisotropic. Typically, this anisotropy is characterised by an ellipse, as shown in Figure 5, whose major axis Y' is along the direction of greatest spatial continuity and whose minor axis X' is perpendicular to this. The direction of the anisotropy is given by the angle θ (or equivalently ($\theta+180°$)) of the principal axis from North. Deposition from a plume is generally an anisotropic phenomenon, with factors such as the constancy of the wind direction and the terrain influencing the shape of the ellipse.

In practice, it can be difficult to establish anisotropy from the data alone, particularly when there are few measurements. In such cases, supporting information such as wind direction can be used to predict the likely direction of maximum continuity. The anisotropy is expressed quantitatively as the range in the direction of minimum continuity divided by the range in the direction of maximum continuity, where the range is defined as the maximum separation for which values may be considered correlated. Data, which are isotropic, will therefore have an anisotropy parameter equal to one, and a highly anisotropic data set will have a parameter close to zero.
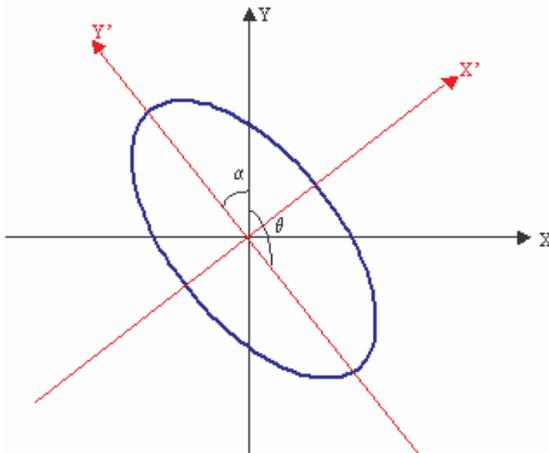
FIGURE 5 Anisotropy ellipse.

3.1.1　　Variograms

Geostatistics is based on an observation or assumption of spatial correlation; that is, measurements taken at locations that are close together are likely to have more similar outcomes than are those taken further apart. Semi-variograms (often simply referred to as 'variograms' in the geostatistical literature, a convention that will be followed in this report) are the principal tool used in geostatistics to quantify this change in correlation

with increasing distance. The variogram may be defined as half the variance of the increment {Z(x) – Z(x+h)} in the random function, as shown in Equation 3.

$$\gamma(x) = \frac{1}{2} E \left\{ (Z(x) - Z(x+h))^2 \right\}$$

(3)

A distinction needs to be made between the empirical or sample variogram, which is plotted for a finite set of sample data and which does not have a value for every possible separation distance, and the model variogram. The model variogram is the result of fitting one or more of a limited number of mathematical functions to the sample variogram, with the models chosen to allow the construction of positive definite kriging matrices. The positive definiteness condition is imposed to ensure that a single stable solution to the kriging equations is obtained.

There are few hard-and-fast rules for selecting which model(s) to use when fitting an empirical variogram. The two most commonly used models are probably the spherical and exponential models, which are both transition models; that is, they reach a maximum or 'sill' value. The spherical model actually reaches its sill value at a separation distance equal to the range, whereas the exponential model approaches its sill asymptotically and has a 'practical range' defined as the separation at which it reaches 95% of the actual sill. If the phenomenon is very continuous, a Gaussian model may be used, which is also a transition model approaching its sill asymptotically. Often these models are used in conjunction with a 'nugget' effect model that represents a discontinuity at the origin (in particular, kriging matrices using the Gaussian model are unstable unless a nugget is included)[*]. From the definition of the variogram, its value for a zero lag[†] must be zero, since in this case the correlation of a measurement with itself is being measured. However, if the data are 'noisy', for example, because of measurement errors, the variogram value for a very small non-zero lag can be quite large. The nugget effect model is equal to zero at the origin and has a constant value for all other distances.

The other common model is the power model. It is distinct from the others discussed here in that it does not reach a sill and its behaviour at the origin depends on the power chosen. Phenomena which are known or assumed to be 2nd order stationary should not be modelled with the power model[‡].

---

[*] The nugget effect represents the unresolveable variation between measurements due to mircoscale processes and measurement error.

[†] 'Lag' is the term used in geostatistics for the separation distance between pairs of locations.

[‡] This model is only positive definite for intrinsic random functions, which are not required to obey the full 2nd order stationarity conditions of Equations 1 and 2. An intrinsic random function is one for which second order stationarity applies only to the *increments* {Z(x+h)-Z(x)} rather than the function itself. Whilst it is true that a second order stationary random function satisfies the intrinsic hypothesis, it is not necessarily true that a random function satisfying the intrinsic hypothesis will be second order stationary (see Wackernagel, 1995 (p36)). Whilst the variogram may be defined for any random function satisfying the intrinsic hypothesis; other measures of spatial continuity (eg the covariance) may not.

Figure 6 illustrates the models discussed above with the power model variogram represented by a linear term directly proportional to the lag (h). The model variograms used in SECTAR calculations were all created using the package Variowin 2.2 (Pannatier, 1996).
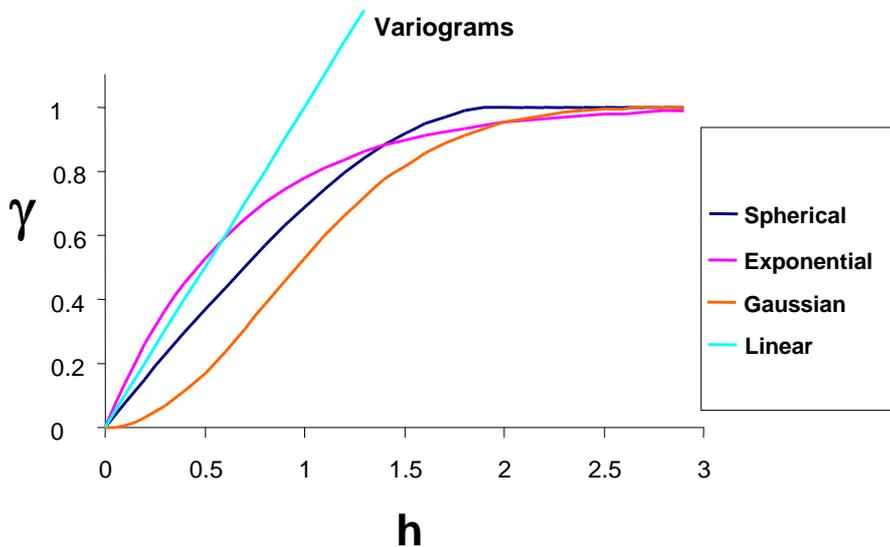
FIGURE 6 Some commonly used variogram models.

3.1.2 Kriging

Kriging is the generic name for a set of techniques that may be used to estimate the value of a variable at locations where it has not been measured. In its most basic form, a kriging estimate is a weighted linear sum of the measured data. The weights are derived from the variogram or covariance model chosen to represent the phenomenon of interest. Kriging therefore provides an estimate that is more sophisticated than simpler techniques, such as nearest neighbour or inverse distance weighting, because it accounts for the spatial structure of the phenomenon in a way that these techniques cannot[*].

Several variations on the basic kriging algorithm exist which may be appropriate in different situations, these are outlined in Sections 3.1.2.1-4. However, before discussing these it is appropriate to introduce two additional pieces of terminology, which are used frequently to describe experiments which compare the predictions of any type of kriging with the available real measurements. 'Cross-validation' refers to an experiment where a single location is selected from a dataset and the data at the remaining locations are used to calculate a kriging estimate at that point. The cross-validation error, defined as the difference between the kriging estimate and the measured value, is then calculated for that location. This procedure is repeated for each location in the dataset in turn, and the mean cross-validation error that results will give an indication of how good the

---

[*] More advanced non-statistical approaches have been shown to be equivalent to kriging. These and other approaches to spatial estimation are discussed in Higgins and Jones (2003).

kriging estimates are. 'Jack-knifing' experiments involve taking a subset from the global dataset and using these measurements to calculate kriging estimates at the non-subset locations. Again, the estimates can be compared with the measured values at each non-subset location, and an average error calculated. Both cross-validation and jack-knifing have been used extensively with the three SECTAR datasets.

3.1.2.1 Simple kriging

This, as the most basic form of kriging, assumes that the measured values are realisations of a stationary random function with a constant mean m. The mean must be known but a variant of the procedure will provide an estimate of the mean over the area (which in turn will reduce to the arithmetic mean of the measurements if they are not spatially correlated). The simple kriging estimator for a stationary random function Z($\mathbf{x}$) is given by Equation 4.

$$Z_{SK}^*(\mathbf{x}) = \sum_{\alpha=1}^{n} w_\alpha(\mathbf{x}_\alpha) Z(\mathbf{x}_\alpha) + \left[1 - \sum_{\alpha=1}^{n} w_\alpha(\mathbf{x}_\alpha)\right] m \qquad (4)$$

In Equation 4 the simple kriging estimate is calculated for location $\mathbf{x}$ using the measurements at the n locations $\mathbf{x}_\alpha$, their mean Z_{sk}^* m, and the kriging weights $w_\alpha(\mathbf{x})$.

In most practical situations, a single mean for an entire area is unlikely to provide an adequate approximation and the technique of ordinary kriging (see below) is to be preferred.

3.1.2.2 Ordinary kriging

This filters the mean from the simple kriging estimator by imposing the condition that the kriging weights $w_\alpha(\mathbf{x})$ sum to 1. This results in the bracketed term in Equation 4 disappearing and creates an estimator that can be used in the more usual practical situation where there are unknown local means, each appropriate to the current local search area. It may be described by the acronym BLUE – Best Linear Unbiased Estimator. It is linear since the estimate is formed from a weighted linear sum of the measured values (within some search radius), and unbiased because it aims to make the average difference between estimated and true values equal to zero, that is, there are no systematic errors which would result in estimates being consistently above or below the true value. In addition to minimising the *average* error, ordinary kriging also aims to minimise the *individual* errors for each estimate. It does this by attempting to minimise the variance of the estimation errors, and in this sense, it is the 'best' estimate. The ordinary kriging estimator for the stationary random function Z($\mathbf{x}$) is given by Equation 5a, which is subject to the constraint of Equation 5b. An example of the solution of this system of equations and the kriging estimate produced for a single location is given in Appendix A.

$$Z^*_{OK}(\mathbf{x}) = \sum_{\alpha=1}^{n} w_\alpha(\mathbf{x}_\alpha) Z(\mathbf{x}_\alpha) \qquad (5a)$$

$$\sum_{\alpha=1}^{n} w_\alpha(\mathbf{x}_\alpha) = 1 \qquad (5b)$$

The optimum size of the local search areas for ordinary kriging is generally dependent on the particular data. There are differing opinions on the subject of what is an appropriate search radius. The number of sample pairs contributing to an experimental variogram is generally small at short lags, increases to a maximum at intermediate separations, and then declines as there are very few pairs of points at large lag separations. Thus, the most reliable data in a variogram plot tend to be at intermediate lag separations. This information can be used when fitting the model variogram to the experimental data but it could be additionally enforced by restricting the search radius used in a kriging calculation. In this case, the search radius would not exceed half the variogram range. However, it is also suggested that a radius even greater than the range could be used (Isaaks and Srivastava, 1989). Investigations using the Bryansk data (Section 5) were carried out to determine the optimal isotropic (circular) search radius. The global dataset containing 1226 points was used for this purpose, so it is clearly not a calculation which could be performed in a realistic accident situation when there are more likely to be a few tens of measurements at most. However, it gives a useful guide as to the 'rule-of thumb', which might be applied when choosing a search radius for other data sets. For this experiment, the technique of cross-validation was used, as described in Section 3.1.2.

The Bryansk global data were modelled with a nugget, a spherical model of range 7.9km and a Gaussian model with range 25.4km. Ordinary kriging was carried out using these models and a selection of isotropic search radii between 1 and 50km. It was found that the mean cross-validation error was the same for search radii between 10km and 50km (-0.023Bq), and that this error increased with further reduction of the search radius. Using a greater search radius led to no further improvement in the estimate. It may be expected that large-scale trend effects (see Section 3.1.2.3) could reduce the reliability of estimates reliant on data at large lags for some datasets. Thus, these results support the idea that it is appropriate to use a search radius approximately equal to the variogram range.

3.1.2.3 Kriging with a trend
In both simple and ordinary kriging, the assumption of *local stationarity* is made; that is, the mean value of the variable over the search area is assumed constant. However, as discussed in Section 3.1 above, in some circumstances a trend is observed in the data such that the mean varies over the search area and it is therefore no longer locally stationary. This idea is illustrated schematically in Figure 7, where random fluctuations take place around a mean that is also changing.
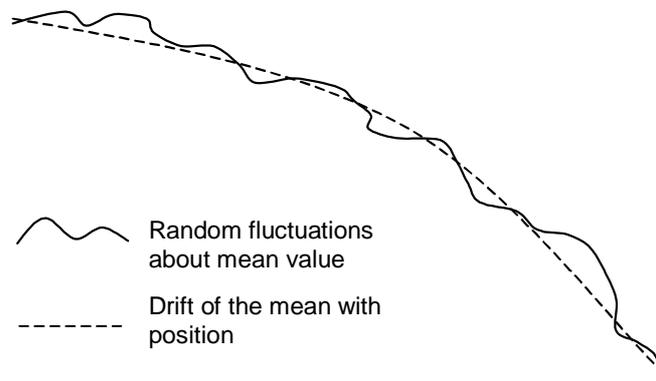
FIGURE 7 Illustration of a non-stationary phenomenon.

Kriging with a trend, or universal kriging, is a variant of ordinary kriging that can incorporate the effect of a trend on the local mean. It does not require prior knowledge of the mean, but does require a model to be supplied for the trend surface. The universal kriging algorithm can generate the trend model by fitting a polynomial function to the local data. Alternatively, the trend can be supplied as an external (secondary) variable[*].

Kriging with a trend uses a random function model that is expressed as the sum of a trend and a residual. The variogram is calculated for the residuals, and used along with the specified trend to krig the original data. The choice of trend is to some extent arbitrary in that the variation can usually be decomposed into trend and residuals in several ways, none of which, of necessity, is clearly favoured by the underlying physics.

When applying a trend as an external variable it is the shape of the trend that affects the estimation by introducing changes to the relative size of nearby residual measurements. This trend surface is rescaled for the local search area surrounding each location where an estimate is required.

A requirement for kriging with a trend is the specification of a model variogram of the residuals from that trend. Whether the trend is defined as a polynomial surface or by means of an external variable that is locally rescaled, this may prove to be difficult. The practical solution adopted is usually to identify a zone in which the trend is less dominant. The experimental variogram is created from measurements in that weak-trend zone. This has the drawback in that it reduces the amount of data on which the variogram is based. It can also be problematic if such a zone cannot be identified.

If the density of data is sufficiently high that it is possible to work with a small search radius, there will be little difference between ordinary kriging and kriging with a trend for estimates within the interior of the estimation area (Journel and Rossi, 1989). This is

---

[*]An alternative to the polynomial approach to the same basic division is to use neural network kriging, (see Higgins and Jones, 2003).

because ordinary kriging re-estimates the local mean for each search area, and can be considered as kriging with a trend where the trend is the constant local mean. In the interior, the local mean is very similar to the trend value for a small search area. When extrapolating beyond the edge of the dataset, the algorithm has to rely increasingly on the trend, whether it is given explicitly when using kriging with a trend or calculated by ordinary kriging as the local mean for the data at the edge. Journel and Rossi recommend that ordinary kriging is used in preference to kriging with a trend if only interpolative estimates are required, unless there is a good physical reason for the phenomenon being modelled to be explicitly decomposed into trend and residuals. Kriging with a trend has been investigated within SECTAR and compared with ordinary kriging using data from Tomsk (see Section 4.2). The approach was justified in this case, as there was a reasonably obvious choice of a trend in the form of a Gaussian plume model representing the dispersion of radioactivity from a source through the atmosphere.

3.1.2.4   Block kriging and change of support
Kriging calculations of the types discussed above result in point estimates. It is often more useful or physically realistic to calculate an estimate of the *average* value of the variable over some area. One way to do this would be to calculate point kriging estimates at many locations within the area of interest and then to use them to calculate an average estimate. However, this becomes computationally intensive when many such 'block' estimates are required, and the problem is better attacked using block kriging, which is a more efficient method for obtaining identical results. Where point kriging uses the covariances between sample points and the point at which an estimate is required, block kriging replaces these with point-to-block covariances.

A closely related topic is that of block support. The term 'support' is used to refer to the volume or area represented by the samples. Taking samples of the same variable at different levels of support can give very different results, with point support measurements typically having a greater variance than those with larger block support, which has a 'smoothing' effect. It is therefore important to consider the physical phenomenon being measured to decide whether samples that have been assigned point locations are really representative of area averages. Conversely, if measured values are to be compared with estimates obtained, for example, from a process model, it is important to ensure that the values being compared share a common level of support (see for example Section 6.6).

3.1.3   Cokriging
Cokriging is an extension of the basic kriging algorithm which allows one or more supplementary variables, which are spatially correlated (or assumed to be correlated) with the primary variable of interest, to be included in the estimation process. This is potentially useful if there are few samples of the quantity of interest but a greater number of measurements of the correlated variable. The development of the method is very similar to kriging with a single variable and derivations can be found in, for example, Chiles and Delfiner (1999). Estimates are calculated from a linear sum of the primary and secondary variables with appropriate weighting factors and suitable

constraints on the weights to ensure an unbiased estimate. Cokriging will only improve on the estimates of ordinary kriging when the primary variable is appreciably undersampled. It has been shown (Goovaerts, 1998) that in the isotopic case, where values of primary and secondary data are available at all measurement locations, mean absolute errors using ordinary kriging are not reduced when cokriging is used, since the relatively weak influence of the secondary data is 'screened' by the primary data at the same location. If more secondary than primary data are available over the estimation area, then a cokriging estimate may be able to provide an improvement over ordinary kriging. However, the respective spatial coverage of the primary and secondary data also needs to be considered (see Section 6.5). The particular requirements to be met and some of the difficulties associated with the technique of cokriging are discussed in more detail in Appendix B.

A variation on the standard approach to cokriging is the much simpler collocated cokriging (Xu *et al*, 1992). In this method, an estimate is calculated by using the primary data within the search radius and a single secondary datum at the required location of the estimate. This simplifies the variogram modelling for cokriging considerably (see Appendix B) and since a secondary datum at the location where an estimate is required will receive much greater weighting than more distant secondary data, including this more distant data is unlikely to result in a very much better estimate. The practical application of the technique is clearly dependent on the organisation of the sampling scheme; given that measurements are currently made by several different organisations, it is quite possible that the measurements of the sample type chosen as secondary data will not cover all the locations where estimates of the primary sample type are required, in which case the collocated cokriging option is ruled out. Some experiments in applying cokriging to the Windscale dataset are discussed in Section 6.5, and they illustrate that one of the principal difficulties in using the technique with post-accident measurements is the choice of the most appropriate secondary data.

3.1.4   Simulations

Geostatistical simulation differs from estimation (ie kriging) in that simulation attempts to model the variability in the surface. Kriging obtains the best estimate value of the quantity of interest at a given location. However, a surface generated in this way is likely to be unrealistically smooth. In simulation, a surface, known as a *realisation,* is generated which has similar variability properties to the real surface but at the expense of accuracy at any specific location – ie its shape, roughness, etc, should resemble the real surface. The difference between the output of the two techniques is illustrated in Figure 8.

Figure 8a is an imaginary cross-section of land with three measurement points. Figure 8b shows the surface which might result if the three points were used to generate an estimate of the cross-section using kriging. The resulting surface has an estimate at every location but it is very smooth and does not really bear much resemblance to the real landscape. Figure 8c is one possible realisation generated by simulation. It is closer in appearance to the real situation although an overall measure of accuracy would indicate that it is a worse estimate than the kriged surface.
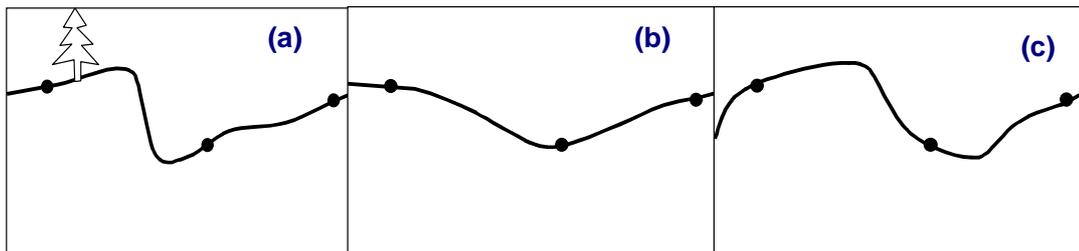
FIGURE 8 Comparison of (b) kriging and (c) simulation estimates based on 3 spot height measurements of a cross-section of land (a). See text for explanation.

Realisations are generated using a random number based algorithm that obtains its information about the spatial structure of the phenomenon from the variogram. The most common use of simulation is to calculate the probability distribution of an event occurring, for example the probability that, if measured, the concentration of radioactivity at some point will exceed a given level. Large numbers of equally probable realisations are generated, each of which is a surface passing through the known measurement points (a simulation, which is subject to this condition, is termed a 'conditional simulation') but differing from the other realisations between the measurement points. At locations far from any measurement points, the values of the different realisations will be more widely distributed than they are closer to measurements, where more information is available, and at the measurement locations themselves the distribution collapses to the single known value (or to a distribution which depends on measurement error). This is illustrated in Figure 9.
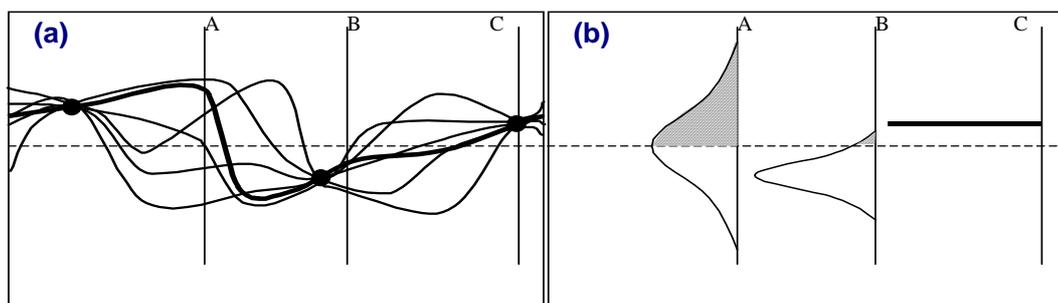
FIGURE 9 Probability distributions from results of simulation using the three measurement points of Figure 8(a). Figure 9(a) shows the actual cross-section as a bold line, along with 5 possible realisations based on the measurements. Probability distributions (see text for explanation) are considered halfway between measurements at A, near a measurement at B, and at a measurement location, C. Figure 9(b) shows how the distributions vary with proximity to a measurement location.

Figure 9a shows the three measurement locations on the cross-section of Figure 8, along with five realisations. The probability distributions are considered at A, B and C where A is midway between two measurement points, B is quite close to a measurement and C is at the location of a measurement. The probability distributions for the simulated value at these locations are shown in Figure 9b, and from these the

probability that the land height exceeds a given value at A, B or C can be estimated. It can be seen that the probability of this height being greater than the height marked by the dotted line is about 50% at A, about 5% at B and 100% at C, because the height at that point has been measured and is known within experimental error.

3.1.4.1   *Sequential Gaussian simulation*

All of the simulations carried out within SECTAR have used the simulation algorithm SGSIM from GSLIB (Deutsch and Journel, 1998). This is an implementation of the sequential Gaussian approach and has been chosen because it is a standard and well-used method suitable for continuous variables. A number of other algorithms exist which may be more appropriate in other circumstances, such as when working with categorical variables, but they will not be considered further in this report. The general theory behind Gaussian simulation will be discussed here and the results of its application to particular datasets presented in Sections 5.2, 6.3 and 6.4. The technique has been used to generate probability contour maps and as an aid in determining where further measurements should be taken, to reduce the uncertainty in contamination estimates.

A sequential simulation is one in which the conditioning data for any location includes not only the measured values but also all previously simulated values within a specified region around the location. When a single quantity is to be simulated, it can be represented as the realisations of a random variable Z_i at N nodes arranged on a grid. The realisations are conditioned, initially by a set of n measurements over the area covered by the grid, to give the conditional cumulative density function (ccdf)[*] of Equation 6.

$$F_N(z_1,\ldots,z_N/(n)) = \Pr\{Z_i \le z_i, i = 1,\ldots,N\}/(n)\}$$ (6)

An N-variate sample (ie a realisation) can be drawn from this ccdf using the univariate ccdfs at each grid node as follows: at the first grid node, draw a value from its univariate ccdf with the n measurement values as conditioning data. Move to the next grid node and draw a value from its univariate ccdf, using the n measurements plus the first simulated value as conditioning data. Continuing this process for every node, the number of conditioning data increases until the Nth node is reached, where the conditioning data set will contain n measurement values plus (N-1) simulated values from the previous nodes. (The order in which the nodes are visited is determined by a random number seed, which may, or may not, be changed if multiple realisations are required[†].)

Sequential simulation therefore requires the determination of the N univariate ccdfs, given by Equation 7.

---

[*] Note not to be confused with a complementary cumulative distribution function (ccdf) commonly used in probabilistic risk analysis studies to present results.

[†] If nodes are visited in the same order for each realisation then kriging weights only need to be calculated once (their values are independent of the sample values) saving considerable CPU time. However, realisations may then be too similar.

$$\Pr\{Z_i \le z_i/(n+i-1)\} \quad i=1,\ldots,N \tag{7}$$

In sequential Gaussian simulation, these ccdfs are all assumed Gaussian and their means and variances are calculated by kriging. The Gaussian simulation algorithm works with the normal score transformed[*] data rather than the raw measurements; the variogram must therefore be modelled for these data with the constraint that the nugget and sill must sum to one. When a simulated normal-score value has been calculated for all nodes, a back-transform is applied to obtain values for the variable of interest.

Simple kriging is usually the most appropriate technique to use when carrying out the simulation of the normal-score transformed data because a stationary random function is assumed in the theoretical development. Deutsch and Journel (1998) give a number of suggestions as to when a non-stationary random function model could be used (including the use of a trend model). For example, if there are sufficient data, using a different stationary model for different areas of the grid, each with its own normal score variogram. Alternatively, a stationary normal score variogram can be used with a non-stationary mean for the random variable which is re-estimated at each location using ordinary kriging. However, they warn that this usually gives poorer results and state that simple kriging should be the 'preferred algorithm for the simulation of continuous variables unless proven inappropriate'.

Within SECTAR, Gaussian simulations have been run for the Bryansk and Windscale datasets. The Tomsk data were considered unsuitable for these experiments, as they are strongly trended, and as discussed above this presents problems for standard simulation algorithms. The Windscale data are also trended due to both the effect of plume dispersion and terrain on deposition and the simulation studies undertaken, discussed in Section 6.3, primarily use ordinary kriging. The consequence of using ordinary kriging is that data values may be spread beyond their range of influence ie the reproduction of the variogram model may be poorer (Deutsch and Journel, 1998). However, it is unlikely that in the early stages of an event there will be sufficient data to consider applying simple kriging over a patchwork of areas, each with its own stationary model. The choice is therefore between applying simple or ordinary kriging using a single stationary model. The quality of the resulting simulations is assessed in Section 6.4 using the techniques of Section 3.1.4.2.

3.1.4.2 *Accuracy and precision of simulations*
The realisations produced by simulations can be used to model the probability distributions of the unknown true values at each point in space. However, considering the theoretical assumptions and practical realities involved in such calculations it is important to have a method of validating the results, analogous to the cross validation and jack-knifing used in the assessment of kriging. The question which needs to be considered is therefore how 'good' the modelled distributions are in relation to the 'true' values. Deutsch (1997) has proposed definitions of accuracy and precision for

---

[*] The normal-score transform converts the histogram of the data to a normal distribution, whilst preserving the quantiles of the original distribution. It therefore does not alter the spatial correlation between values in the distribution.

simulations, which quantify this problem. The definitions are based on the 'symmetric probability intervals' of the modelled distributions. A probability interval is 'symmetric' if a random variable taken from the distribution has an equal probability of being below the interval as above it. The 'width' of a probability interval is defined as the probability that a random variable taken from the distribution is inside that interval. For example, for a symmetric probability interval of width 0.5, a random variable has probabilities of 25%, 50% and 25% of being below, in and above the interval respectively.

Deutsch (1997) defines accuracy and precision as follows. In these definitions, p refers to the width of the symmetric probability interval:

"A probability distribution is 'accurate' if the fraction of true values falling in the p-interval exceeds p for all p in [0,1]".

"The 'precision' of an accurate probability distribution is measured by the closeness of the fraction of true values to p for all p in [0,1]".

For each fraction, p, it is expected that p of the true values will lie in the symmetric probability interval of width p, eg, 50% of the true values should lie in the 50%-interval of their associated distribution. A simulation is deemed accurate if there are at least as many true values as expected for every possible width of the probability interval.

Precision is a measure of how closely an accurate simulation resembles the ideal situation, where *exactly* p of the true values lie in the probability interval of width p. It is possible, for example, to imagine distributions in which 100% of the true values all lie in the 80% symmetric probability interval of their simulated distribution. In this case, the simulation is suggesting that there is a high likelihood of values outside the range observed, ie the simulated distributions are too wide. Thus, the simulation is accurate but not precise. A more precise simulation would include all the true values within say, the 95% symmetric probability interval[*]. A good model has both high accuracy and high precision. Deutsch (1997) quantifies these ideas, giving expressions which may be evaluated for the accuracy, precision and goodness (a related quantity which gives an overall measure taking account of both accuracy and precision). A model, which is accurate but not precise, is being overly pessimistic about the degree of uncertainty.

The accuracy and precision of a chosen model can be checked using either cross-validation or jack-knifing (see Section 3.1.2). The general approach is to use a large number of realisations to produce a large number of simulated values at each location where a real measurement value is known. From these values, the conditional cumulative density function (ccdf) may be plotted for each location. For each p interval, an indicator function $\xi(p)$ can be defined at each location such that $\xi(p)=1$ when the true value at that location falls within the p interval, and $\xi(p)=0$ when it does not. For each interval p such that $0 \leq p \leq 1$, the average indicator over all locations, $\overline{\xi(p)}$ is then the proportion of locations where the true value falls within the p interval. The average indicators can then be plotted against the probability interval to obtain a graph, from which the accuracy and precision of the simulation may be assessed, see Figure 10. An

---

[*] In terms that are more familiar, an accurate estimate will contain the true value within the error bounds and a precise one will have narrow error bounds.

accurate simulation will generate a plot in which $\overline{\xi(p)} \geq p$ for all $0 \leq p \leq 1$. The precision is judged by how close the graph of $\overline{\xi(p)}$ against p is to the 45° line on the graph, ie the line corresponding to the situation when the proportion of true values falling within each p interval is actually equal to p. Deutsch favours the graphical accuracy plot over the use of the numerical measures of accuracy, precision and goodness for evaluating the accuracy and precision of a simulation (Deutsch, 2000). Plots of this form, produced for the Windscale simulation trials, are shown and discussed in Section 6.4.
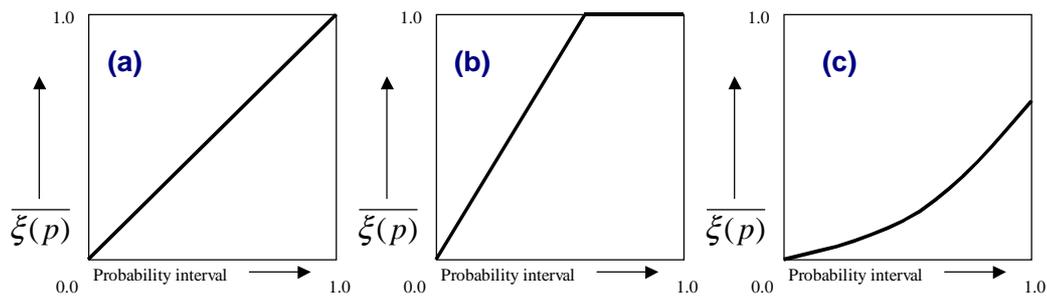
FIGURE 10 Features of accuracy plots. (a) A plot indicating good accuracy and precision. (b) A plot indicating that the ccdf is too wide, it is accurate but not precise. (c) A plot indicating that the distribution is too narrow, it is not accurate.

3.2 Bayesian methods

Bayesian methods cover the entire gamut of statistical modelling, differentiating themselves from the classical frequentist approach by introducing the concept of a subjective probability that exists without the need to postulate an infinite number of trials. The subjective probability can, through the use of Bayes' theorem, be changed under the influence of evidence ie as trials progress the Bayesian estimate will become more similar to the results of frequentist trials. There remain arcane arguments between frequentists, Bayesians and philosophers as to the nature of probability, and the validity of Bayesian and other approaches. However, Bayes' theorem and the tools of the trade are not generally in dispute and the approach provides many practical advantages.

In the context of data assimilation after an accident, Bayesian analysis can provide a method of evolving the predictions of models to take account of measurement information as it arrives. Thus, the prior expectations of critical model parameters are updated through the influence of measurements[*], and the uncertainty in predictions is

---

[*] In the statistical description of a process, the model parameters characterising the phenomena are not single numbers but random variables. Measurements provide information to modify the distributions of these variables and consequently the expected value of the model (see Section 3.2.1).

reduced as they become increasingly based on measurements rather than on the generic model[*] and subjective information used initially.

The implementation of the Bayesian approach under SECTAR has been carried out in collaboration with Professor O'Hagan and colleagues in the statistics department of Sheffield University, and further technical details may be found in Kennedy et al (2002). A simple Gaussian plume atmospheric dispersion program (Clarke, 1979) was modified to take account of measurement data when estimating deposition values. The resulting 'Bayesian' dispersion program goes further than adjusting the model parameter values in a calibration process and includes consideration of *model inadequacy* which gives a measure of how well the predictions of the process model compare with the observed data (Section 3.2.2). The main features of this Bayesian program are described in the following sections.

3.2.1 Model parameter uncertainty

The predictions of the Bayesian program before any data are collected are based entirely on the output from a process model (in this case the Gaussian dispersion model). In reality, few if any, of the parameter values required by this process model will be known exactly. However, in practice, most of the values are known approximately which allows a tractable result, with uncertainty, to be produced by assuming that only a small number of important parameters are uncertain for any particular scenario.

In this implementation, the source term Q and the deposition velocity V_G of the dispersion model are assumed to be unknown but can be calibrated against field data when these become available. For generality, it is assumed that there is an unknown level B of background contamination, which is added to the predicted deposition level as a constant. (This background may be due to deposition, which occurred earlier in the accident if there are multiple releases or even deposition from a previous accident). The *effective* process model output may then be written in a form consistent with the information from measurements as

$$\eta(x, y, \boldsymbol{\theta}) = B + \eta(x, y, Q, V_G) \tag{8}$$

where $\eta(.)$ is the output of the process model for a single plume (the contribution from additional plumes could be added if required).

Beliefs about the 3 parameters (B, Q, V_G) are described using a probability distribution. The parameters Q and V_G have been singled out because they are both difficult to quantify and will have a significant effect on the predictions.

Let θ = (log(Q), log(V_G), log(B))[*] denote the vector of the 3 parameters whose values are unknown. It is then assumed *a priori* that beliefs about θ can be approximated using a Normal distribution, which is written as:

---

[*] The Bayesian formulation employed has the option of minimising the influence of an inadequate model if interpolation using only the available measurement data will provide a better result (see Section 3.2.2).

$$\theta \sim N(\theta_m, \theta_v) \tag{9}$$

Here θ_m is the prior expectation of θ and θ_v is the prior covariance matrix for θ. Expert knowledge can be included in the model by specifying suitable values for θ_m and θ_v. For example, to express vague information about θ, large diagonal elements (variances) for θ_v would be used. All other parameters for the plume model are fixed at suitable values for the scenario under consideration. Of the many parameters that are included in the formulation of a Gaussian dispersion model, these were chosen as representing the most critical and event specific. Additional uncertain parameters could be added to the formulation, but as more parameters that are uncertain are included, more measurement information will be required to narrow the uncertainty ranges. However, as the chosen parameters will only tend to scale the deposition it may be appropriate in future studies to examine uncertainty in other parameters of the process model. The general method is not restricted to this or any other particular choice of deposition model. Different process models could easily be used to provide a prior estimate, but simple models are favoured for several reasons:

a   The process model is only used as a starting point, and observational data are used increasingly to modify its predictions. In principle, if there were enough data the predictions would be effectively independent of the process model.

b   Complex process models typically include more unknown parameters. Specification of prior probability distributions for these parameters can be a problem. The number of unknown parameters also has the most significant impact on computation time for the Bayesian analysis. This is because with p unknown parameters, a number of p-dimensional numerical integrations are required.

c   Complex process models may themselves require a significant amount of computing time. Although the influence of the process model is reduced as more data are obtained, large numbers of model evaluations are required in order to learn about the model inadequacy (see Section 3.2.2) and update the probability distribution for the process model parameters. Slow complex process models will therefore slow the Bayesian analysis significantly.

With only two uncertain model parameters and the Gaussian plume model, the method can be used to plot contours of predicted deposition and variance maps within a few minutes for each new set of observations.

---

[*]Each element of θ is strictly positive, and it is convenient to work with the log scales, particularly when integrating with respect to these parameters.

3.2.2 Model inadequacy

In most situations, the process model will not predict the actual deposition with a great
deal of accuracy, even if the 'correct' input parameters are known[*]. Thus, in addition to
the parameter uncertainty described above, the prediction must be corrected for 'model
inadequacy'. Model inadequacy is defined to be the difference between the true values
at a particular location and the prediction of the process model at that location when the
best parameter values are used to evaluate the process model. The best values of the
model parameters are initially unknown but may be solved for approximately by
calibrating the process model against the available measurement data obtained for that
particular event. Before the results of any measurements are available, generic
parameters are used which represent the best fit of the model to a variety of data.
These parameters may be estimated based on previously published data, if it is
considered suitable, or by expert judgement.

Let $d(x, y)$ denote the true deposition at co-ordinates (x, y) and let $\eta(x, y, \theta)$ denote the
corresponding process model output, where θ is defined as above. The assimilation
model assumes that for a *known* value of $\theta = \theta$, determined, for example, by calibration
or expert judgement, the deposition can be written as:

$$\text{Log }\{d(x,y)\} = \rho \log\{\eta(x,y,\theta)\} + I(x,y,\theta) \tag{10}$$

where $I(x, y, \theta)$ is the model inadequacy function and ρ is a regression parameter to be
estimated from the data. The regression parameter allows for a systematic discrepancy
between the process model representation and reality. Thus, the process model may
predict the deposition to be of broadly the same form ie to have areas of relatively high
or low deposition similar to that observed but to vary either too much or too little
depending on whether ρ is below or above 1. It is important to note that this definition
involves known values for the parameters in θ. Thus, this term forms one part of a
hierarchical model, in which θ and $[d(x, y) \mid \theta]^{\dagger}$ are evaluated separately. It is more
natural to think about $[d(x, y) \mid \theta]$ in this way and deal with uncertainty about θ as a
second level of modelling.

3.2.3 Modelling the inadequacy function

The inadequacy function $I(x, y, \theta)$ is modelled as a *Gaussian process*, for reasons of
simplicity, tractability and flexibility. Modelling the function in this way allows the
correlation between its values at different locations to be represented in one of several
simple alternative forms. The properties of the Gaussian model are as follows:

For any (x, y, θ), the expected value of $I(x, y, \theta)$ is a constant β and for any pair of points
(x, y) and (x', y') the covariance function is:

---

[*] 'Correct' can take on a variety of meanings depending on the application. It may, in some cases,
refer to the parameters that minimise the least squares error between the model and measurements.
However, this might result in a very poor representation in key areas, in which case parameters may, if
possible, be chosen to improve predictions in these key areas.
[†] The probability of deposition d given a particular θ

$$\mathrm{Cov}\{I(x, y, \theta),\, I(x', y', \theta)\} = \sigma^2 \mathrm{c}\{(x, y),\, (x', y')\}$$

(11)

where, for example, the correlation function $c((x, y), (x', y'))$ can be assumed to have the following *product Gaussian* form.

$$\mathrm{c}\{(x, y),\, (x', y')\} = \exp(-b_x |x - x'|^2)\exp(-b_y |y - y'|^2)$$

(12)

Thus, for any set of co-ordinates $(x_1, y_1, \theta)\ldots(x_n, y_n, \theta)$, the corresponding values of $I(\cdot)$ have a multivariate Normal distribution. This property is true of any Gaussian process by definition.

The choice of the covariance function is important, as this characterises the belief in the smoothness properties of the deposition (the quantity of interest in the assimilation model). It defines the nature of the local influence on the observed data points and acts in an analogous way to the variogram used in geostatistics. In the case of the product Gaussian correlation form, for example, the constants, b_x, b_y, and σ^2 are parameters of the Bayesian model, whose values are estimated from the data. For example, large values of b_x and b_y indicate large local variations in the deposition value, so that a measurement can only provide information about neighbouring values within a small region[*]. The product form for $c(\cdot,\cdot)$, through the adoption of different values for b_x and b_y, can also be used to characterise differences in the extent of correlation in the downwind and crosswind directions following deposition from the atmosphere in an analogous way to the use of an anisotropy ellipse in geostatistics.

The inadequacy term acts to 'correct' the predictions of the process model if the measurements indicate that it provides a very poor representation of the phenomenon under study. Under these circumstances, the action of the inadequacy term ensures that the contribution of the process model to the final estimate will be very small and the technique will produce a form of kriging estimate using the measurement data[†]. The ability of the technique to handle inadequacy is particularly important when models are used in the initial assessment of consequences of actual accidents. In this case, only simple models that require only the small amount of available input data can be used, and it is therefore unlikely that they will be able to represent the physical situation accurately. For example, Gaussian dispersion models reproduce the general trend of atmospheric dispersion over a short distance but are not well suited to handling changes in wind direction, rainfall or complex terrain. This will be demonstrated in Section 6.7 where the results of using the Bayesian program with Windscale deposition data are presented. The program has also been used with data from Tomsk, which is well described by a Gaussian plume model. These results are presented in Section 4.3.

---

[*] Note the correlation structure assumed is generally simpler than the form used in geostatistics, which may be composed of several structures with distinct ranges of influence.
[†] The calibration parameter ρ (see Equation 10) will reduce the influence of the process model to let the inadequacy term improve predictions.

3.3 Bayesian assimilation

After the initial (prior) distributions for the calibration parameters and the model inadequacy term have been established, the Bayesian program is able to update them, as more deposition measurements become available. This allows the *posterior distribution* of the deposition to be calculated, that is, the probability density function of deposition estimates, which have been revised given the additional information from the measurements. The Bayesian prediction takes account of the uncertainties in the measurements, the model and in the calibration parameters; it also estimates the scaling parameter and the model inadequacy term. This process of refining prior distributions and inadequacy estimates as more data arrives is a form of Bayesian assimilation.

4 APPLICATION OF TECHNIQUES TO TOMSK

The Tomsk aerial gamma survey data was the first to be used within the SECTAR project. As discussed in Section 2.1, this is a set of 812 measurements of ^{106}Ru made 5 months after the release[*]. However, many of these 812 measurements were below the limit of detection (LoD) of the airborne detector used, thus making them less useful. Comparisons with the results of model calculations were generally made with and without the aid of these censored measurements, which when included were all given the value of the lower LoD (3700 Bq m^{-2}). The deposition pattern appears to be remarkably similar to the idealised form produced by a simple Gaussian dispersion model (Clarke, 1979). However, although this classic plume shape can be discerned, the plume predicted by such a model (see Figure 2) is narrower than the one observed, even after adjustment of the model parameters (as advised by Shershakov *et al*., 1995). Another feature of the Tomsk data is that the detector used in the aerial survey has a 'footprint', ie the area on the ground contributing to the detector signal, which is of the order 100-150m across. Thus, the measurements have larger support (see Section 3.1.2.4) than ground-based measurements. Larger support will smooth out small-scale variations in the data, so the results will be averaged across the footprint. This has little impact on the calculations carried out for Tomsk, because of the lack of data with different support. However, it would need to be considered if ground based data were available and the two sources of information were to be used together. Support is addressed in the analysis of the multiple data types available in the Windscale data set (see Section 6).

These complications aside, the data from Tomsk can be described as 'simple' in that they are as ideally represented, as any real measurements can be, by the predictions of a simple atmospheric dispersion model. This however, makes any demonstration of the benefits of statistical techniques using this data into a difficult test case, as they must perform particularly well to improve upon the predictions of a simple process model.

---

[*] Other radionuclides were measured but did not provide distinctive additional information and were not used.

The first experiments were designed to compare the effectiveness of ordinary kriging and kriging with a trend using datasets of various sizes. Twelve subsets, four each of 20, 40 and 60 points, were drawn at random from the global dataset, excluding censored values. The aim of the experiments was to use these subsets to estimate the non-sampled values in the global data set, including censored values. Excluding the censored data from the global data set introduced a bias towards high values that would then be reflected in a similar bias in the subsets drawn. Although not exactly characteristic of aerial survey sampling this was likely to be a realistic representation of ground survey sampling where, in many cases, interest would be drawn to areas of expected high deposition. The subset sizes were chosen to represent the paucity of data available in the early stages of an accident. From a rigorous geostatistical perspective, the smaller subsets are at (or beyond) the extreme limit of acceptability, because it is difficult to form a variogram with such a limited supply of data. This is an important issue for the practical application of geostatistics to accident data as any statistical technique chosen should be able to provide a timely analysis using the measurements available.

Variograms were modelled for each of the subsets and for the global dataset (including censored values set to 3700 Bq m^{-2}). This process became more difficult as the size of the subset decreased since in this situation the empirical variograms had a tendency to be erratic and to differ greatly from the global dataset variogram. For example, although there was clearly anisotropy in the global dataset, with the principal axis at 30°, this could be lost or completely distorted in the subsets and could only be inferred from external information such as wind direction. The kriging routine chosen, ktb3d, (Deutsch and Journel, 1998) allowed the specification of a search ellipse which reflected the anisotropy of the datasets.

The comparison between the subsets was performed using a jack-knifing procedure[*]. For the Tomsk data, these locations were divided into the measurements which were above the LoD and those which were below it (and were all set to 3700 Bq m^{-2}). The jack-knifing was performed on the two groups of data separately because it was expected that the technique would perform differently for the censored measurements, which were on the edge of the plume and therefore among the least well positioned for estimation. Additionally, because the true values for these data were unknown, it was expected that estimations of kriging error at their locations would also be less reliable.

The subsets were also jack-knifed in the same way using the variogram model derived from the global data set, although obviously this information would not be available at the time of an accident. These calculations were carried out to investigate the effect of a better variogram model on the results, if such a model could be obtained or derived.

Finally the subsets were kriged to generate estimates on a regular grid; to give deposition patterns that could be assessed visually. Departures from the expected

---

[*]Jack-knifing as previously introduced is when subset data are used to generate kriging estimates at the remaining, non-subset, locations in the global dataset. The kriging error, defined as the difference between the kriging estimate and the measured value, is then calculated at these non-subset locations.

pattern, such as unexpectedly high or low activity regions in the plume, were qualitatively evaluated.

The experiments described above were performed using both ordinary kriging and kriging with a trend. It is clear from Figure 11 that there is a trend in the data – its marked plume shape. On this basis, ordinary kriging, with its assumption of a constant local mean, was not expected to be the most appropriate technique. It was nevertheless included since, as it only uses the raw data and a variogram, it provides a simple baseline against which other techniques can be compared. The ordinary kriging experiments are introduced in Section 4.1 but detailed discussion of the results is deferred until they can be compared with results from kriging with a trend in Section 4.2.
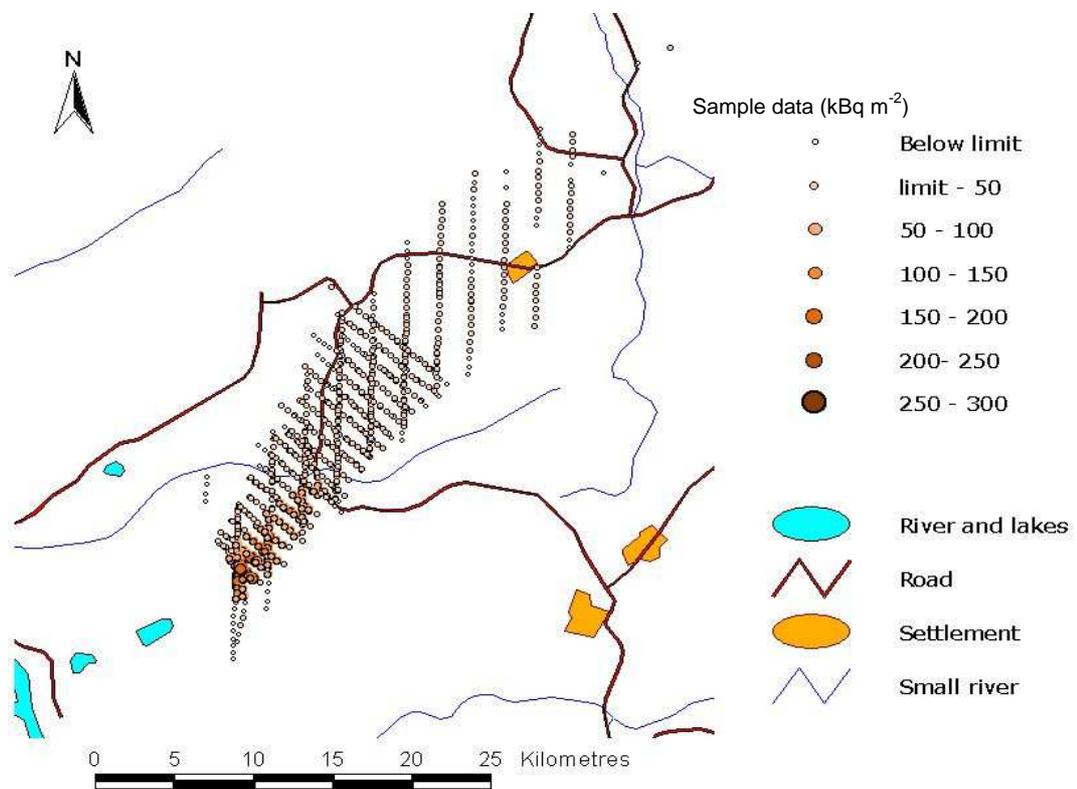
FIGURE 11 Tomsk global data set.

4.1   Ordinary kriging

Trial runs using ordinary kriging with a short search radius (2-4 km, the approximate range of the variograms) produced patchy surfaces, particularly when using the smaller subsets. Increasing the search radius produced more continuous surfaces but gave greater errors and problems, particularly at the edge of the estimation area. This was attributed to the mean over the larger search area being more like the global than the local mean, which would reduce the chance of ordinary kriging being able to perform well. Ordinary kriging re-estimates the local mean for each search area and would therefore require a very small search area if it was to produce good estimates using

data which had a strong trend, ie, a rapidly changing mean. A strategy of using a large search radius but restricting the maximum number of points used to produce a kriging estimate at a location to a small number was adopted. This had the effect of only including distant points in an estimate if there were insufficient data close to the required location. This balanced the conflicting demands of minimising the error and maximising the estimation area.

If all of the Tomsk data are used, ordinary kriging will largely cope with the trend at interior locations. This is because the measurements are at a sufficient density for the change in the mean that occurs between data points to be low (see Section 4.2). Figure 12 shows the experimental and modelled variograms for the global dataset in the major and minor anisotropy directions, and the ordinary kriged estimated surface.
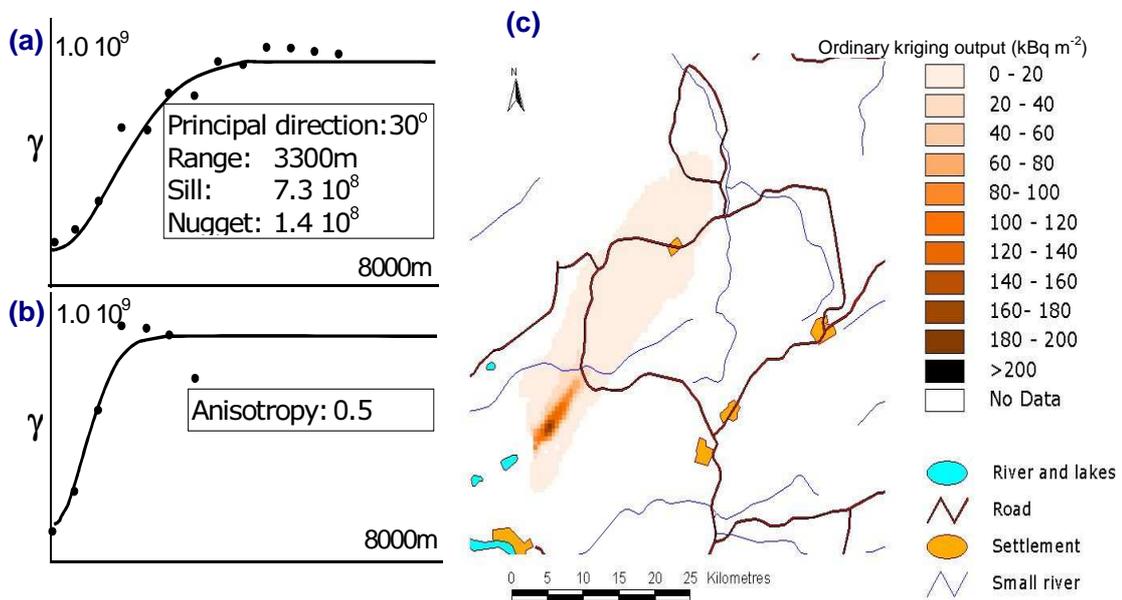
FIGURE 12 Tomsk global dataset variograms in (a) major and (b) minor anisotropy directions, with (c) ordinary kriging estimates made using this variogram model.

Figure 13 shows the variograms and kriged surfaces for example subsets, of 20 and 40 points respectively. All variograms were modelled using a Gaussian model with a nugget. It was felt that the Gaussian model might be an appropriate choice given that the dominant underlying process is often represented deterministically using a Gaussian dispersion model.

Although all of the deposition patterns for the subset plumes show a general pattern of having high values in the middle of the estimated area and low values at the edge, the edges themselves are very 'noisy'. There are patches on the edge where there are high values in otherwise low areas and vice versa. This effect was attributed to the less than ideal spatial configuration of the subset sample points. This meant that when estimating locations at the edge, the search radius had to extend to its maximum to include

enough sample points to generate an estimate. Under these conditions, the kriging algorithm is using a mean, which is not local to the edge and more like the local mean of the interior[*].
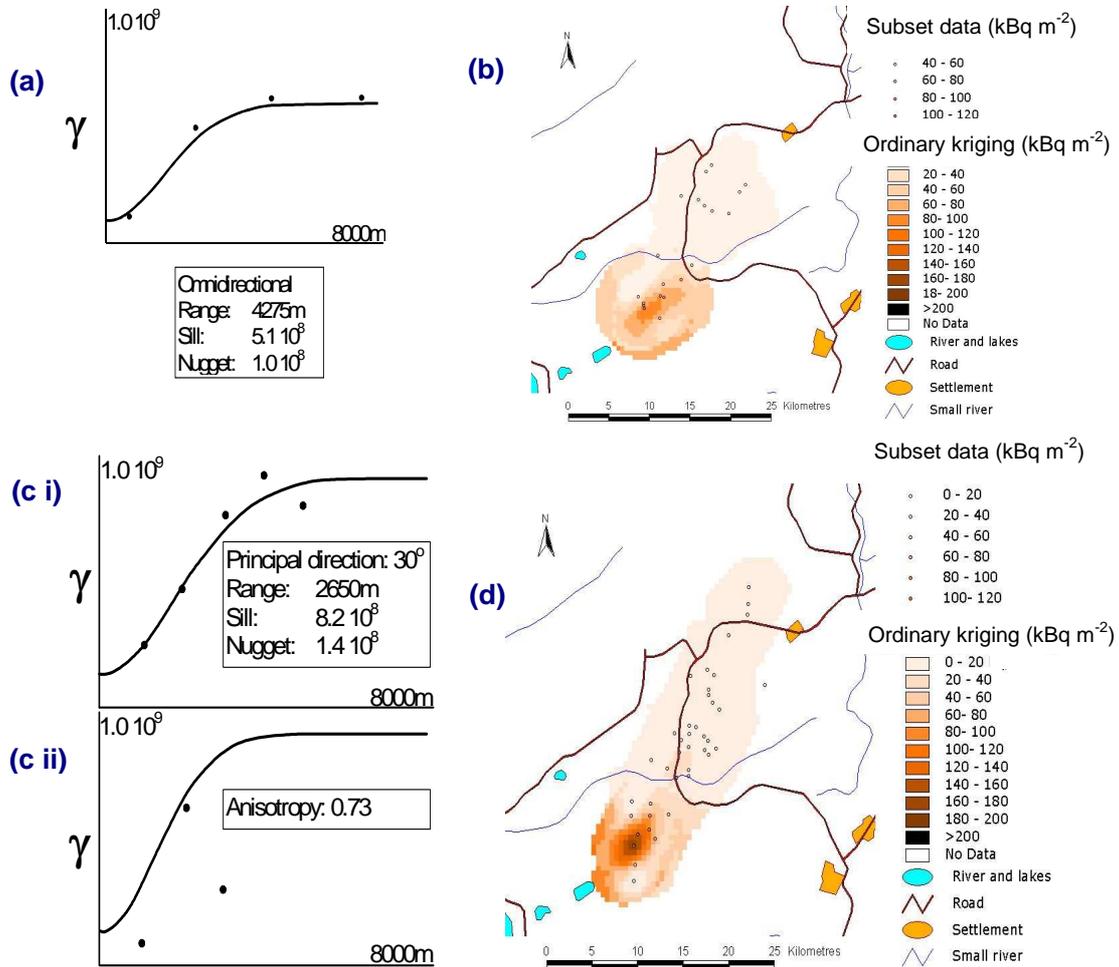
FIGURE 13 Variograms and ordinary kriging results for Tomsk subsets. Part (a) shows the omnidirectional variogram for a 20 point subset, and (b) the kriging results. An omnidirectional variogram is used for the 20 point sample as it is too small to determine the anisotropy. Variograms in the major (c i) and minor (c ii) anisotropy directions for the 40 point subset are illustrated in (c), with the corresponding kriging output in (d).

The edge effects were most pronounced when an isotropic variogram model and search radius had to be used, for the 20-point samples. This observation is most likely due to

---

[*] These will be referred to as Council Food Intervention Levels (CFILs).

the small sample size, as the dependence of the estimate on the precise parameters of the variogram is known to be weak (Wackernagel, 1995)[*].

These features are not surprising given the nature of the Tomsk data, which, as discussed above, does not really satisfy the assumptions of ordinary kriging. It was concluded that little that could be done to improve on the ordinary kriging estimates. The technique has the advantage of being simple to apply and therefore relatively quick but it is not appropriate for small, highly trended datasets or when it is important to make estimates near the boundary of the sample data.

4.2　　　　Kriging with a trend

Trials using kriging with a trend are expected to show an improvement over the ordinary kriging results of Section 4.1. The technique can account for the non-stationarity of the data within the local search area in addition to local re-estimation of the mean provided by ordinary kriging. The trend component of the random function model was supplied by the plume shaped deposition profile generated by a simple Gaussian atmospheric dispersion model (Clarke, 1979). The predicted plume of deposition from such a model is shown in Figure 2 and repeated for convenience in Figure 14. Figure 14 appears to show the model output extending further than the observed deposition, but this is because the model predictions for the outer region of plume deposition are below the limit of detection (LoD). It was mentioned in Section 2.1 that the model used was refined to suit the environmental conditions of the accident. In particular, the deposition velocity and ground roughness length used were both greater than the generic default values recommended for the UK (Clarke, 1979). Much of the Tomsk area is covered by coniferous forest, whereas the default parameter values apply to open countryside. The defaults are appropriate for use in the early stages of an accident in the UK or elsewhere, if more site-specific information is not available. Should more refined estimates from a Gaussian dispersion model be required later, as in this case, the model can be re-run with site and event specific parameters. When applied to the Tomsk data, the use of default parameters in the model results in a deposition pattern similar to that of the refined model but with less material deposited within tens of kilometres of the site. The discussion of Section 4.3 illustrates how model parameters could be adjusted dynamically as new measurement information arrives.

The variograms for the residual global data and subsets were modelled in a similar fashion to that undertaken previously when considering the use of ordinary kriging. However, in this case, the highly continuous nature of the Gaussian variogram model near the origin was not appropriate for modelling the residual noise. The spherical model was therefore chosen (however see Figure 19). When modelling the global variogram, shown in Figure 15(a), measurements located in the highest deposition area near to the origin of the release were excluded. This was the area where the trend was

---

[*] What counts more than the details of the variogram is the type of continuity assumed for the regionalised variable (see Section 3.1) and the stationarity hypothesis associated with the random function (see Section 3.1.1).

strongest and obtaining appropriate residual information the most difficult. However, for the subsets this procedure would result in the further reduction of an already small number of points, so the entire subset was used.
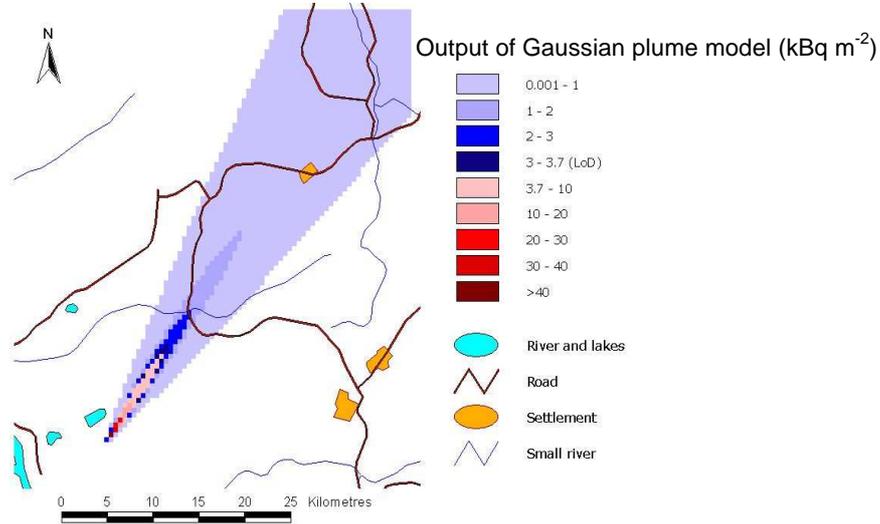
FIGURE 14 Output from a simple Gaussian plume model, using estimated source term and default parameters with the colour coding indicating where measured values are likely to be above the Limit of detection (LoD).
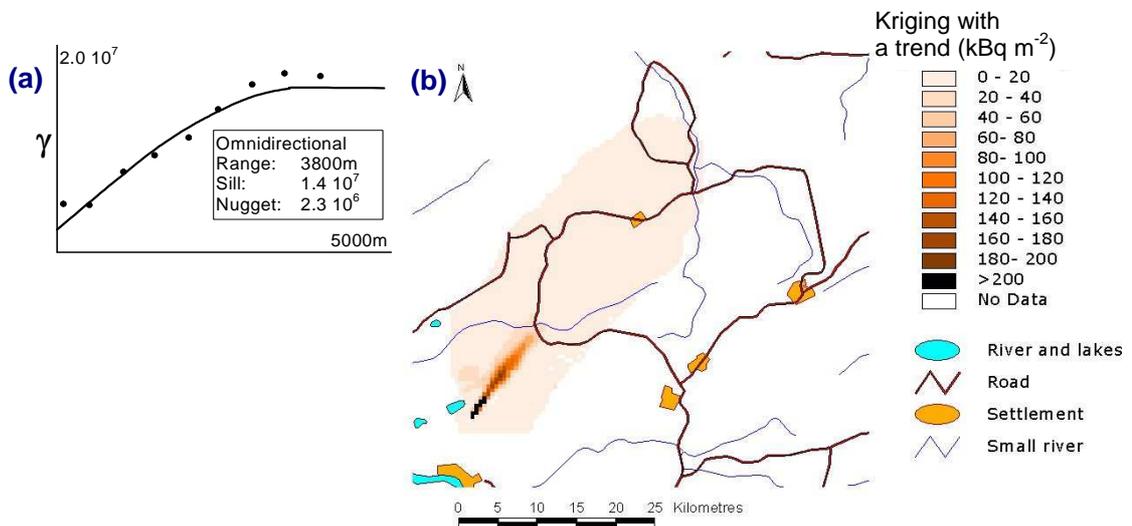
FIGURE 15 (a) Omnidirectional variogram, and (b) surface kriged with a trend for the Tomsk global dataset.

As with ordinary kriging, the jack-knifing was performed for both the global and the subset variogram models. The results, for the global data and the subsets used in the ordinary kriging experiments, are shown in Figures 15 and 16 respectively. The surface in Figure 15 derived from the global data appears to be very similar to that derived using ordinary kriging in Figure 12. This, as discussed previously, was expected since both

estimates use all 812 measurements. However, the area covered by the kriging with a trend estimates is larger than that for ordinary kriging, reflecting the larger search radius that could be used.
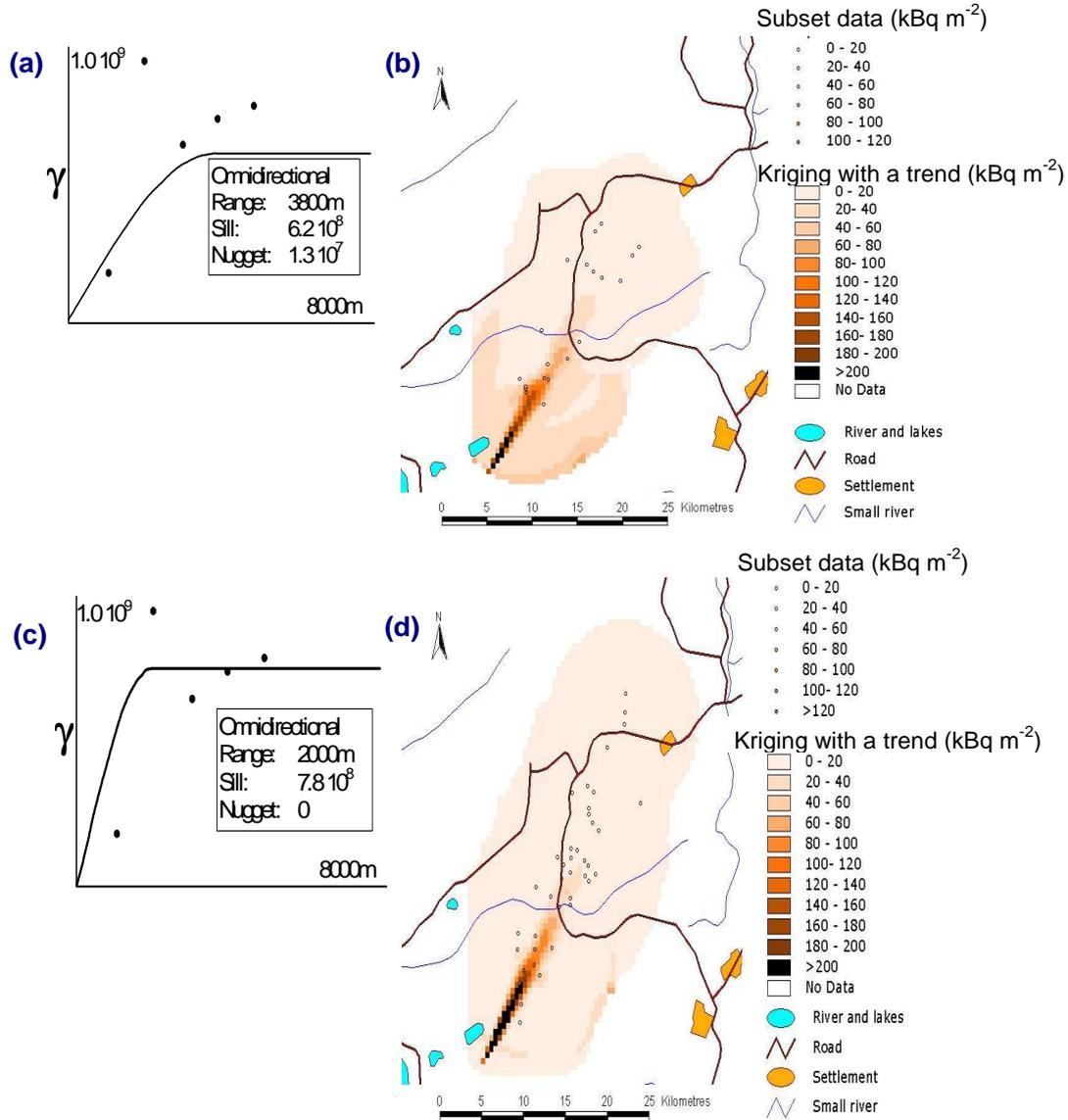
FIGURE 16 Kriging with a trend on subsets of Tomsk data. (a) and (b) show the omnidirectional variogram and kriged surface for a 20-point subset. (c) and (d) show the same for a 40-point subset.

The Root Mean Square (RMS) errors from the jack-knifing of all subsets used in the trials are shown in Figure 17. The results are given for jack-knifing against remaining measurements in the global data set which were above the LoD, and separately for jack-knifing against those which were censored and had been set equal to the lower LoD. These results are further separated into those obtained using the subset variograms and those from using the global variogram.

Considering first the ordinary kriging results, the expected result that larger subsets give smaller kriging error is not apparent except for the uncensored measurements. These measurements tend to be located in the interior of the plume, see Figure 1 and the discussion in Section 2.1. This illustrates that it is the spatial distribution of locations rather than the absolute number of measurements that is important for kriging with small samples. Consequently, ordinary kriging shows particularly poor performance in estimating censored data values, located, as they are, on the periphery of the sample measurements. The RMS errors are greater than for the uncensored data, and when the individual errors are examined, it is found that the technique consistently over-estimates the contamination at these locations. The overestimation is such that the censored data are estimated to have values greater than the lower of the limits of detection (3700 Bq m^{-2}).
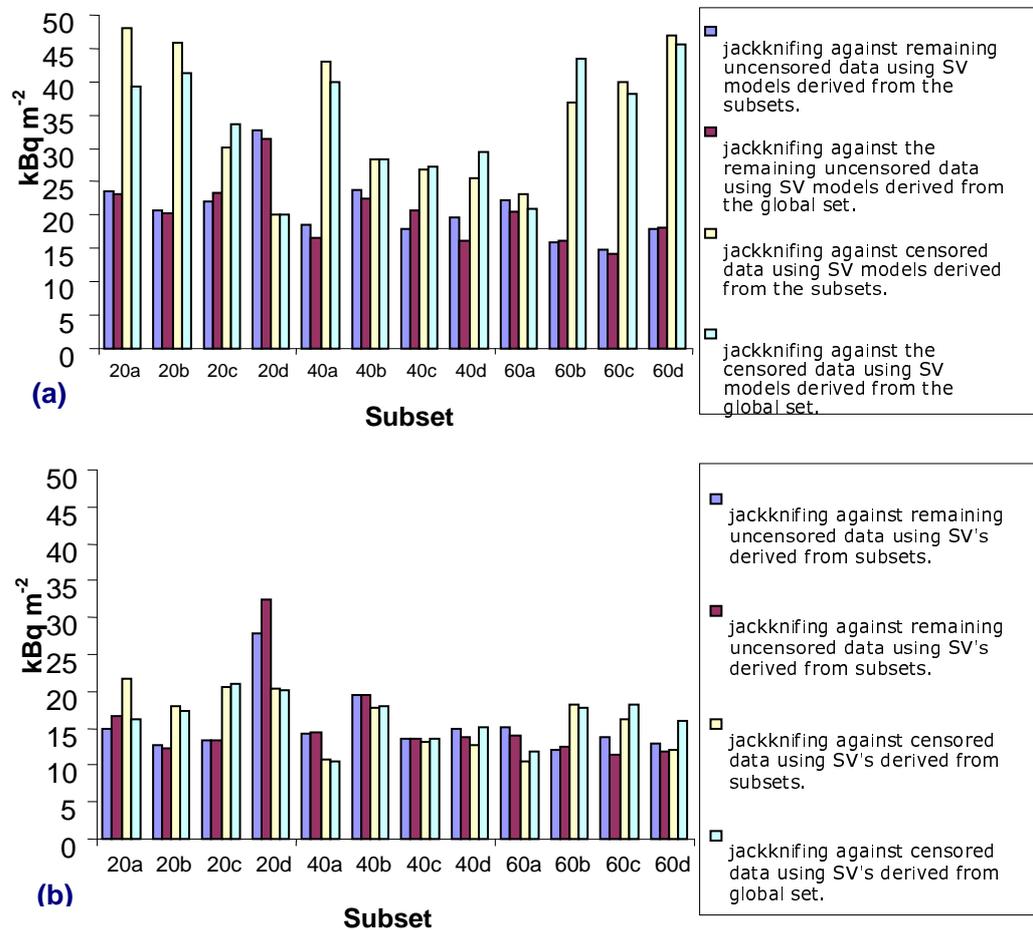
FIGURE 17 Barcharts indicating relative errors for (a) ordinary kriging and (b) kriging with an external trend, on subsets of 20, 40 and 60 points selected from the Tomsk dataset. Results are shown for four different subsets of each size[*].

---

[*] In the legends, variogram is abbreviated as 'SV', in reference of the correct usage of Semi-Variogram (see Section 3.1.1).

In comparison, kriging with a trend almost invariably gives a better estimate than ordinary kriging. Thus, although the separation of the trend from the residual to create the required variograms is a difficult, and to some extent an arbitrary procedure, it does not appear to have a prohibitively detrimental effect on the estimates made. Kriging with a trend appears to estimate the censored measurements as well as the uncensored. Although an examination of the individual errors shows that censored measurements are generally over-estimated, this is by less than occurs when using ordinary kriging, and some estimates of the censored data are below the lower limit of detection. In a sense the algorithm locally fitted the surface shape as defined by the model to the available data within the search radius and kriged the resulting residuals to obtain local estimates.

The definition of residual variograms for kriging with a trend proved difficult not only because of the strong trend but because the samples were collected by helicopter. The samples were collected in four sets of passes, two sets in a north-south direction and two sets taken across the plume (ie northwest to southeast). The subsets were sampled randomly, so samples that appear to be close in space may have been taken on completely separate passes of the helicopter. The height of the helicopter has a big impact on the limit of detection and potentially on the reading for any given level of deposition. Increasing the height of the helicopter not only leads to greater attenuation of the signal from the ground but averages that signal over a greater ground area. Thus, it could be argued that the correct residual variogram to use would be pure nugget. This position is supported by the plot of measurements from different passes along one transect as shown in Figure 18 where the deposition at particular locations was recorded by different passes. The intra-pass variability may be less than the inter-pass variability shown in Figure 18 but the overall expectation must be that the data are likely to have a reasonable amount of superimposed noise.

Neither ordinary kriging nor kriging with a trend errors were significantly reduced by using the global rather than the subset variogram. This suggests either that the global variogram was not significantly better than the subset variograms, or that the dependence of the kriging estimate on the precise parameters of the variogram is weak. It has been shown (Chiles and Delfiner, 1999) that the latter statement is true. An imperfectly specified variogram will not cause the computed kriging estimates to differ greatly from the optimal estimates, as long as the basic models chosen reflect the general characteristics of the data and the behaviour of the variogram near the origin is correct. This is further demonstrated in Figure 19, which compares kriging, including the use of alternative variogram models, and Bayesian estimates.
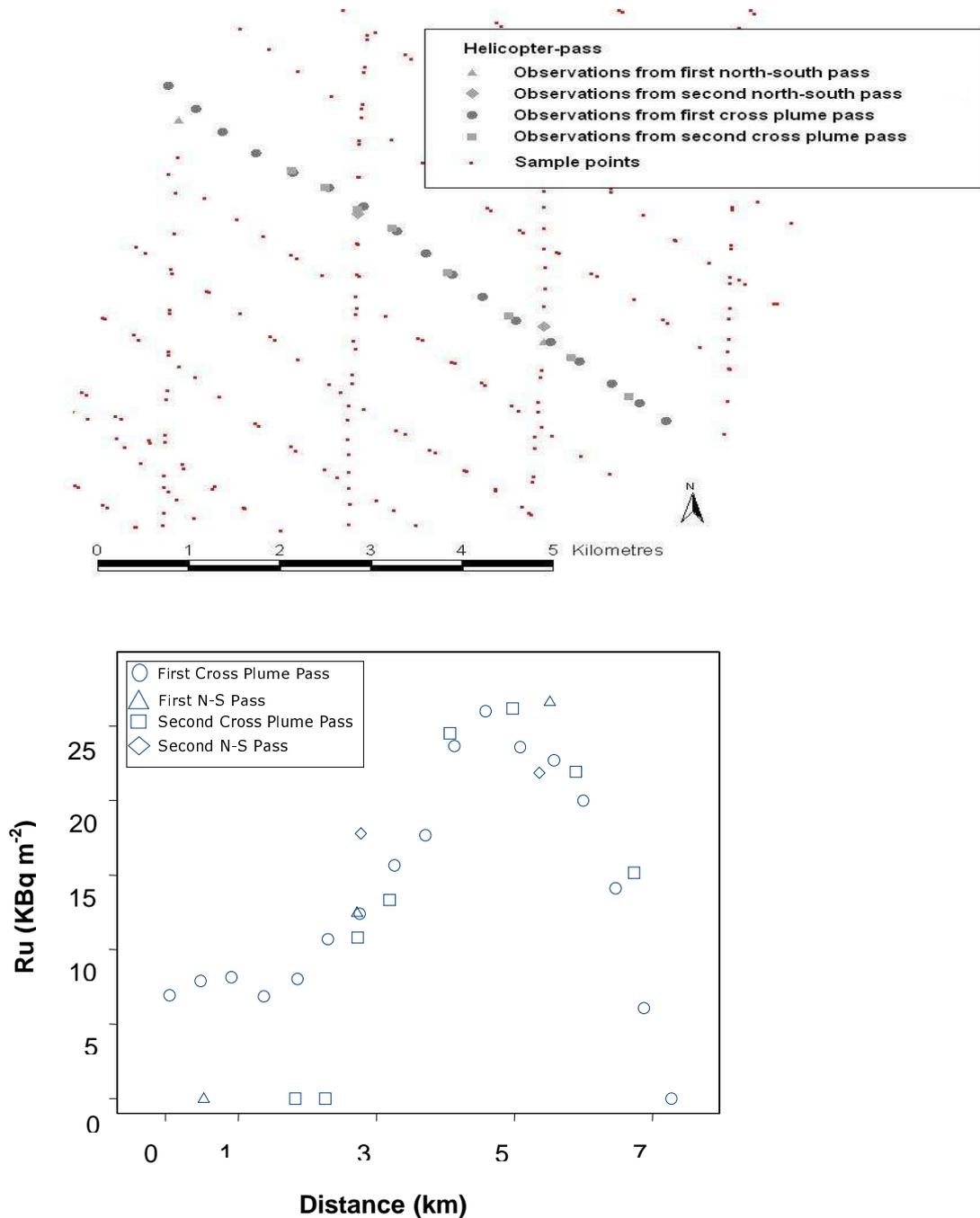
FIGURE 18 Pattern of aerial monitoring at Tomsk and the effect on observed deposition. The points in the graph correspond to measurements in the transect marked on the map, with distance measured along the transect line. It can be seen from the graph that deposition measurements made at the same location vary depending on the pass in which they were made. This variation is principally due to the variation in helicopter height from one pass to another.

4.3   Bayesian

The Bayesian assimilation method developed for SECTAR by Kennedy and O'Hagan (Kennedy *et al*, 2002) and outlined in Section 3.2 has been tested using subsets of the deposition data from the Tomsk-7 accident. Figure 19 shows a comparison of the Bayesian and kriging approaches for a selection of the Tomsk subsets introduced in Section 4.1. The figure shows that the Bayesian and kriging techniques have a broadly similar performance when the RMS error is calculated using the non-censored data found in the interior of the plume. However, if only the censored data (found at the edge of the plume) are considered, the kriging results can be seen to over estimate. Ordinary kriging is particularly bad in this respect as it relies entirely on a local average concentration, which due to a lack of data is determined by values near to the centre of the plume.

Figure 19 also shows the error in the estimated deposition predicted by a Gaussian dispersion model that used the source term and other model parameters determined by Shershakov *et al* (1995) that they thought best represented the particular features of the release. This fitting included the consideration of factors not considered in these tests of interpolation methods such as the ground roughness length. In addition to the refined fit of Shershakov *et al* (1995) the results of fitting the HPA-RPD default Gaussian dispersion model to the data using the respective subset results are shown. Model parameters are set to the standard values used by HPA-RPD and the source term adjusted using least squares to achieve the lowest RMS error between model predictions and measurements. The Bayesian model slightly outperforms this simple approach when the comparison is against the censored data. However, the reverse is true, although again not by very much when only non-censored data are considered. The Bayesian method is not an exact interpolator like kriging ie the estimated deposition at a sub-sample location will not be equal to the measured value supplied. However, the main reason for the relatively poor performance in the plume centre is the use of log values in the Bayesian formulation, which increases the importance of the low values found at the plume edge when applying the fitting procedure (see Section 3.2). The particular Bayesian implementation employed assumes that only two of the Gaussian dispersion model parameters are uncertain, namely the source term and the deposition velocity. It is therefore to be expected that the results will be similar to a simple fitting procedure when a Gaussian dispersion model is a good fit to the data.

Bayesian calculations were also carried out for evolving data sets ie where a subset of 40 samples includes the 20 sample subset. This showed that the predictions were consistent, in the sense that as more data become available, the RMS error is reduced further.
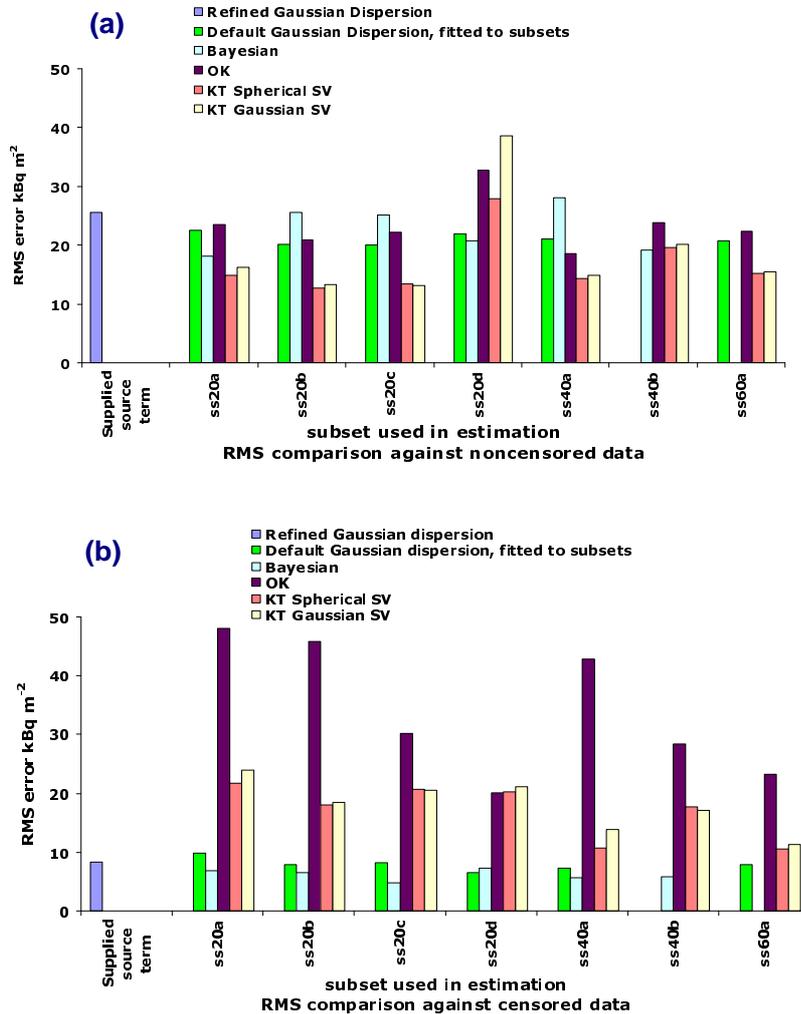
FIGURE 19 Comparison of root mean square errors produced by a Bayesian model, ordinary kriging (OK), and kriging with an external trend (KT), using the Tomsk data which were (a) above the limit of detection and (b) censored.

5   APPLICATION OF TECHNIQUES TO BRYANSK

The Bryansk data are well suited to the technique of ordinary kriging, as demonstrated in Section 5.1, and to geostatistical techniques in general. The data are approximately stationary, with only a slight drift in the deposition recorded when moving from west to east or north to south.

5.1   Ordinary kriging

Ordinary kriging may be expected to perform better for data that do not have a pronounced trend. To test this idea, ordinary kriging was applied to the Bryansk data. A similar procedure to that adopted for Tomsk was applied, with nine subsets being drawn at random from the 1226 measurements in the global data set, three each of 20, 40 and 60 points. Jack-knifing was then carried out and the RMS relative kriging errors[*] calculated for locations in the complement of each subset. This quantity could then be compared with the same quantity calculated for the uncensored Tomsk data. The results are illustrated in Figure 20 and they confirm the hypothesis that ordinary kriging should perform better on the Bryansk data than it does on the strongly trended Tomsk data.

Neither the Tomsk nor the Bryansk data show an obvious decline in the relative error as the sample size used in the estimates increases. In the case of Bryansk, this is likely to be due to the relatively small number of additional locations used in increasing the subset size from 20 to 60. The change from using the smallest to the largest subset only increases the proportion of the global data used from approximately 1.5% to 5%. However, for the smaller Tomsk dataset the lack of improvement in the estimates when the sample size is increased is much more likely to be caused by the continued dominance of edge effects. Sample 60d, for example, has the greatest relative error of all the Tomsk subsets, because it contains a large number of clustered measurements and therefore many more edge areas. Sample locations within a cluster will also have an increased likelihood of being effectively 'redundant' because of their proximity to other measurements.
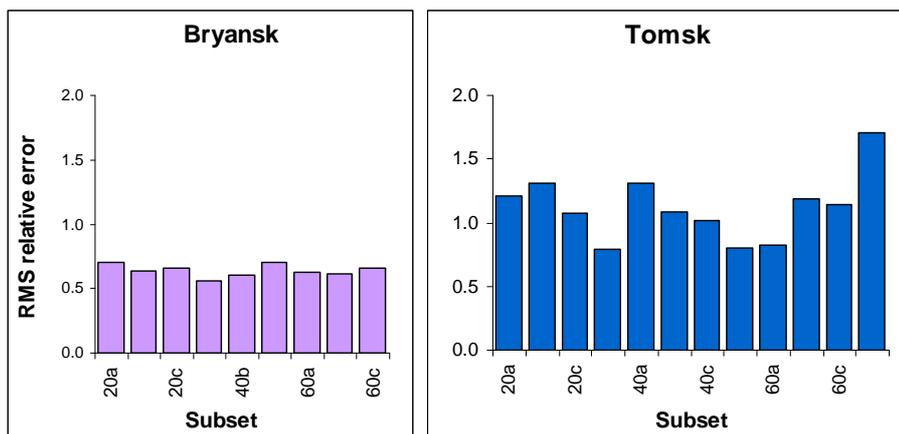
FIGURE 20 Comparison of the RMS relative error in jack-knife estimates found in Bryansk and Tomsk trials using a range of subset sizes.

---

[*] The relative error was defined as the difference between the estimated value and the measured value, divided by the measured value.

5.2 Simulations

Simulations performed using a sample of 40 randomly selected measurements from the complete data set of 1226 measurements were used to plot the contours shown in Figure 21. The values illustrated for each location are the 95th percentile from the distribution of estimates at that location. The contours delimit areas with estimated contamination below the contour value. They show the location and maximum likely extent of areas where, for example, restrictions could be considered. As an alternative, contour maps indicating the probability of exceeding a selected threshold could be shown. Further investigations into the uses of simulation are described in Section 6.3.
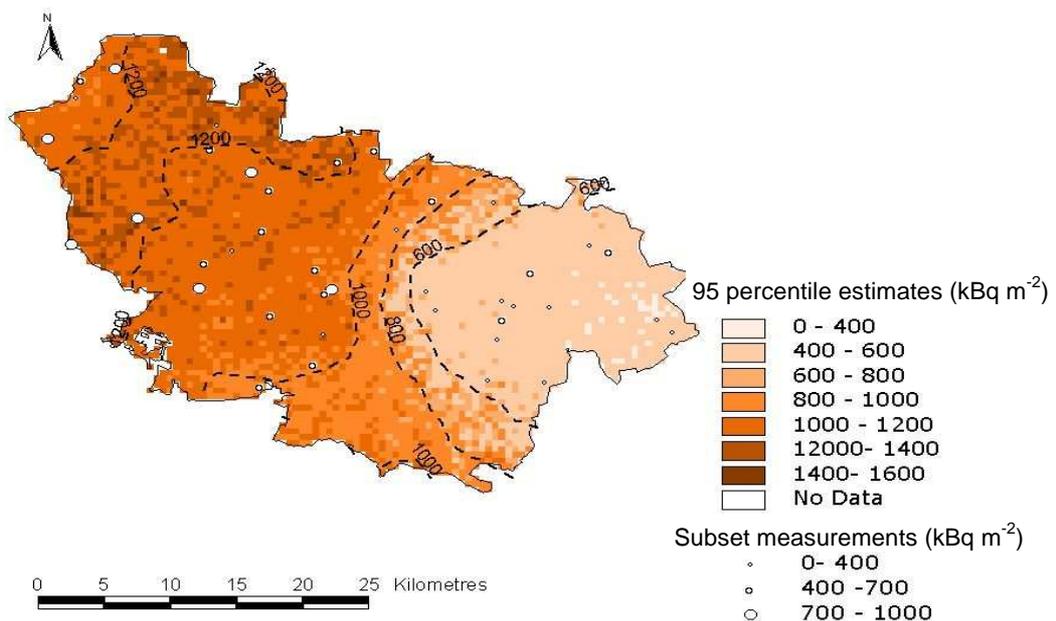
FIGURE 21 Contours showing the 95th percentile estimates of the contamination for Bryansk.

6 APPLICATION OF TECHNIQUES TO WINDSCALE

The available Windscale data differs from those discussed in the preceding sections in containing measurements of a number of different radionuclides in several different sample types, taken over a period of approximately six weeks. To a certain extent, this allows more flexibility in choosing which data to analyse, however, additional restrictions are imposed by the more complicated nature of the release. A comprehensive account of the Windscale fire has been given by Arnold (1992); the discussion here will concentrate on the features of the accident which are most relevant to the geostatistical analysis of the data.

The fire itself was burning for approximately 2 days, having started on 10th October 1957 and finally being extinguished on the afternoon of 12th October following the

application of water for 30 hours. There is some uncertainty about the duration of the release, although it is assumed to have begun on 10th October when air samplers on the site detected beta activity 10 times greater than normal levels (Dunster *et al*, 1958). Arnold (1992) states that by 13:30 on 10th October there was a 'marked increase' in radioactivity in the stack of Pile No. 1. It has been suggested (Chamberlain, 1959) that there was 'serious emission' from 16:00 on 10th October to 11:00 on 11th October, with two peak release periods, the first, of unspecified duration, at about midnight on 10th October and the second starting when water was first applied at 09:00 on 11th October and lasting for approximately two hours. Other complicating features of the Windscale accident were the varying wind direction and speed whilst the release was ongoing, and the Cumbrian terrain which includes both coastal plain and hilly areas. This led to the pattern of deposition being very different from that which would be obtained if the release was modelled using an estimated source term with a simple Gaussian atmospheric dispersion model. Clarke (1974) analysed the Windscale accident using the WEERIE code, which used an estimated fission product inventory and a precursor of the current Gaussian dispersion model (Clarke, 1979) to predict ^{131}I deposition. This obtained a reasonable agreement with a small subset of measurements along a relatively narrow band out to a distance of 50km from the source. Over the full area that we are attempting to fit, such a simple model cannot provide an adequate representation of the overall deposition pattern. A simple Gaussian plume model does not account for the effect of hills, a major feature of the Windscale area, on the dispersion of radioactivity. In addition, it is not valid over large distances or for the complex weather conditions prevailing at the time of the accident. Results obtained using a simple program based on the Gaussian dispersion model (Clarke, 1979) are shown in Figure 22.

Assuming a source term of 1000 TBq ^{131}I (typical estimated source terms in the literature are 700-800 TBq, see for example Chamberlain (1981)), a Gaussian dispersion model (Clarke, 1979) was used to estimate the deposition at locations where actual measurements had been made on or before 22nd October, some 10 days after the release ended. Only those locations less than 5km from the estimated plume centre line[*] were considered, and the maximum distance downwind was approximately 30km. A 24 hour release was assumed and the measured and estimated depositions are all decay corrected to 22nd October using a 3.6 day effective half life for ^{131}I on grass calculated using the HPA-RPD FARMLAND model (Brown and Simmonds, 1995). The discrepancies between the modelled and measured deposition confirm that a simple dispersion model does not provide an adequate representation of the Windscale release.

It could be argued that the comparison of Figure 22 is not entirely fair. A simple dispersion model is much better at estimating the peak deposition than the location of the peak, even in ideal conditions. However, it would also be unreasonable to adjust the

---

[*] The direction chosen for the plume axis was 146°, as given in Chamberlain (1959). Clarke (1974) also makes use of this reference for the plume direction, and reports the wind direction as 'NW' (ie 135°). Chamberlain and Dunster (1958) use 145°. The choice here is therefore consistent with previous work on this data.

centre line assumed for the plume at each downwind distance to optimise the fit to the measurement at that distance.
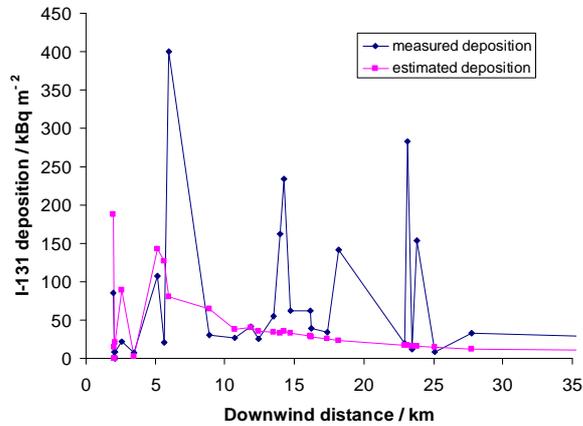
FIGURE 22 Comparison of Gaussian dispersion model estimates for deposition to grass with measurements (lines shown for clarity). The model results are discontinuous because the estimates are made at the measurement locations and these are not generally on the (assumed) plume centre line. All locations are within 5km of the centre line.

6.1 The milk ban and ordinary kriging

The most serious environmental consequence of the Windscale fire was the contamination of milk with ^{131}I. This resulted in the imposition of a ban on the sale of milk for human consumption over an area of 520 square kilometres around the site. This ban was imposed in two stages; the first stage banned milk from farms in the immediate vicinity of the site with effect from 12th October and had an extent of 207 square kilometres. This area was then increased on the morning of 15th October to its final extent. The criterion for the ban was an activity concentration of at least 0.1 μCi l^{-1} (3700 Bq l^{-1})[*].

If such a release were considered in the context of current emergency arrangements, the Food Standards Agency would issue precautionary food safety advice from an early stage of the accident. They would then be required to consider restrictions under the Food and Environmental Protection Act (FEPA) 1985 on the sale of food. This process is expected to take at least 24 hours from the start of any emergency. For a large event such as Windscale the specification of the restrictions is likely to take longer and is unlikely to be definitive until sometime after the release has stopped. The monitoring of milk after Windscale did not begin in earnest until the release stopped. However, it took several days from then to determine the extent of the ban. It was therefore thought appropriate to take advantage of the information available from the Windscale database

---

[*] At the time of the Windscale accident, activity was measured in Ci, and the milk ban criterion was also in these units. Diagrams in this section will show measurements in the original units or in multiples of the milk ban activity level.

to test whether the application of geostatistical techniques could have helped specify the extent of the ban sooner. The initial approach considered was to concentrate on measurements made within the 48 hour period after the fire (ie. on or before 13th October). The ^{131}I in milk measurements made during this time could then be used to re-derive the extent of the area banned in 1957 using measurements made on or before 14th October.

A preliminary analysis of the data revealed that measurements were made at 37 different locations on 13th October. Measurements had been made at 3 locations on 12th October, but these were excluded as there was a possibility that they had been taken while the release was ongoing. In any case, these locations were re-sampled on 13th October. While 37 is a rather small data set for geostatistical purposes, the work on Tomsk and Bryansk had showed that it was feasible to model variograms based on samples of this size. However, the sampling undertaken in 1957 extended further from the site over the days immediately following the accident. Unfortunately, the earliest measurements were all taken close to the site and it was therefore clearly impossible to re-derive the large milk ban area using an interpolative technique such as kriging. The available sample data simply did not cover the area over which an estimate was required. Inclusion of data from 14th October provided an additional 36 measurement locations, which then created a data set that covered the area over which estimates were required. Obviously using the full 73 measurements in this data set would not be an improvement on the methods used in 1957. To create a suitable data set for a geostatistical analysis, four samples of 37 measurements (the same number actually measured on 13th October) were drawn at random from this global set. It was considered reasonable that, had the sampling programme been arranged differently, all 37 measurements could have been made on 13th October, ie within 48 hours of the release[*]. Variograms were modelled for each sample, as illustrated for one example in Figure 23.

Two of the sample data sets showed clear anisotropy, whereas for the other two (including the one illustrated in Figure 23) it was difficult to determine whether they were anisotropic. This is a consequence of the small sample size, with one or two points having a major effect on the overall anisotropy of the sample. However, the global data set[†] of 73 measurements was clearly anisotropic, with the principal direction being determined by both the wind direction, which was from the North-West for the main release period, and by the terrain. These factors combined to direct the plume along the coastal plain, and consequently the highest deposition values tend to be located along this direction. The principal direction was taken as 125° based on the direction of maximum continuity indicated by the variogram surface. The direction of the plume centre line has been estimated (Chamberlain and Dunster, 1958) as 145°, which is in reasonable agreement with the direction assumed here. On this basis, it was decided to model each of the samples with anisotropy in this direction. Ordinary kriging was then

---

[*] At later stages, many more measurements were being made per day, with the number of locations at which samples were taken reaching a peak of 228 on 22nd October.

[†] The global data set would obviously not be available at this stage after an accident; however, it would still be possible to suggest an anisotropy direction based on the physical characteristics of the plume and local terrain.

carried out using each of the samples to calculate estimates on a 0.5km grid. The search radius in each case was taken as the range of the appropriate subset variogram. Convex hulls[*] were then fitted around those measurements which had been estimated as exceeding the 1957 ban limit, as an approximation to the area which would be predicted using ordinary kriging. It was found that three of the subsets resulted in estimated ban areas which were in very good agreement with the area designated in 1957 (see Figure 24). Although they were all larger than the original area, it should be remembered that they were estimated using half the number of measurements. In any case, it is more acceptable to reduce a conservative estimate for a restricted area, as more data become available than it is to increase one that is found too small. It can be seen from Figure 24 that one of the samples (sample 3) results in an especially large overestimate of the ban area. From the kriging surface, the cause of this can be identified as a combination of the configuration of points in the sample and a large search radius. In practice, when deciding a restriction area this effect could be corrected for by further sampling in specific areas. The kriging variance is a simple measure of the uncertainty in estimates resulting from the spatial arrangement of sample data, and can be used to identify those areas where additional measurements may be beneficial. These results highlight the importance of a well-designed sampling scheme, particularly when relatively few measurements can be made

(a)

(b)
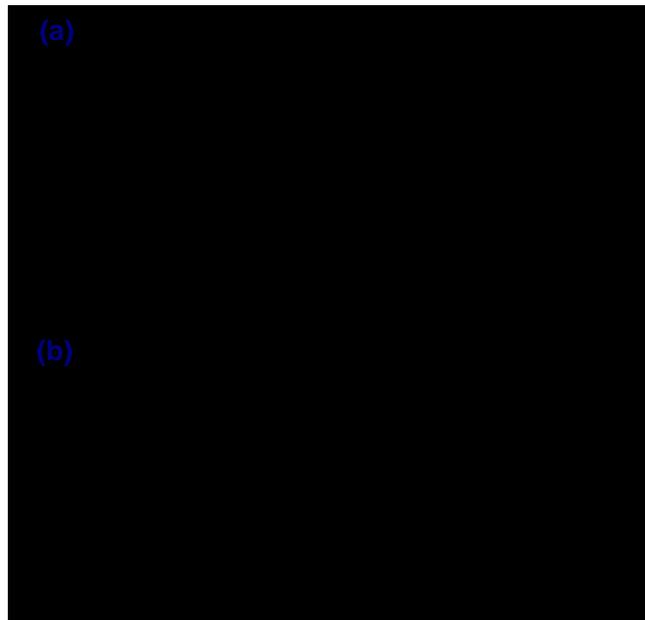
FIGURE 23 Example variogram model along (a) major anisotropy axis, and (b) minor anisotropy axis, for a 37 point random sample of the available Windscale milk measurement data.

.

---

[*] A convex hull is a polygon having the property that a chord drawn between any pair of points on its perimeter always remains within the polygon. The convex hulls in this case therefore enclosed the maximum area which could be restricted using the $0.1\mu Ci \ l^{-1}$ limit.
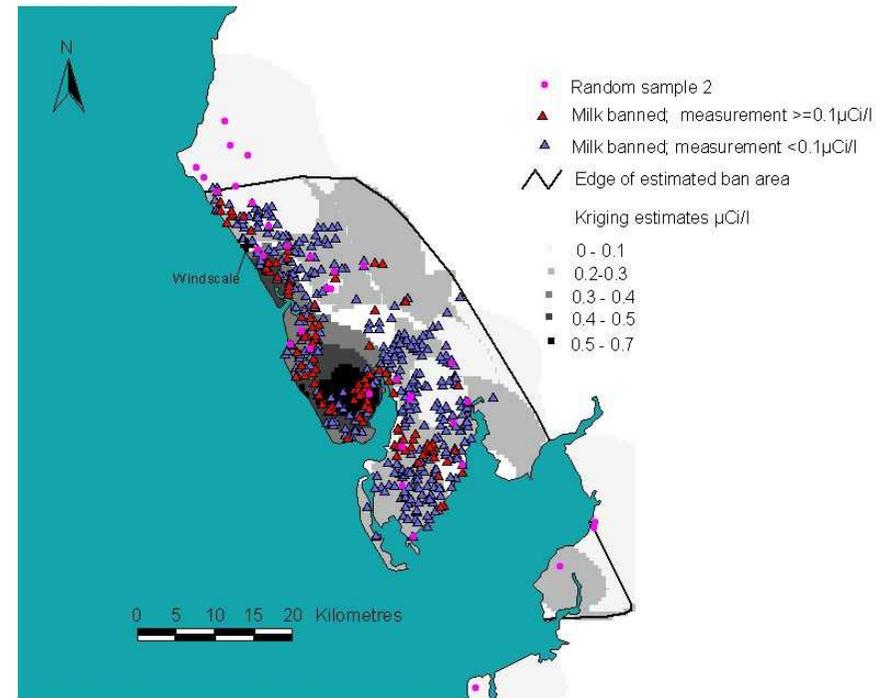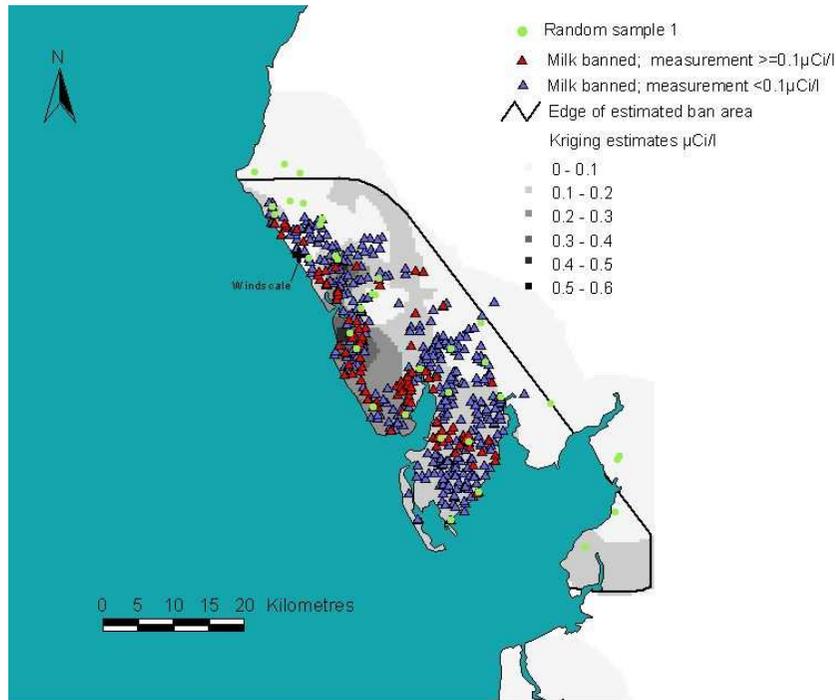
FIGURE 24a Comparison of areas predicted to exceed the milk ban criterion used in 1957, 0.1μCi l⁻¹ (3700Bq l⁻¹), with the location of restricted farms derived using random samples 1 and 2 of four randomly located selections of milk measurements.
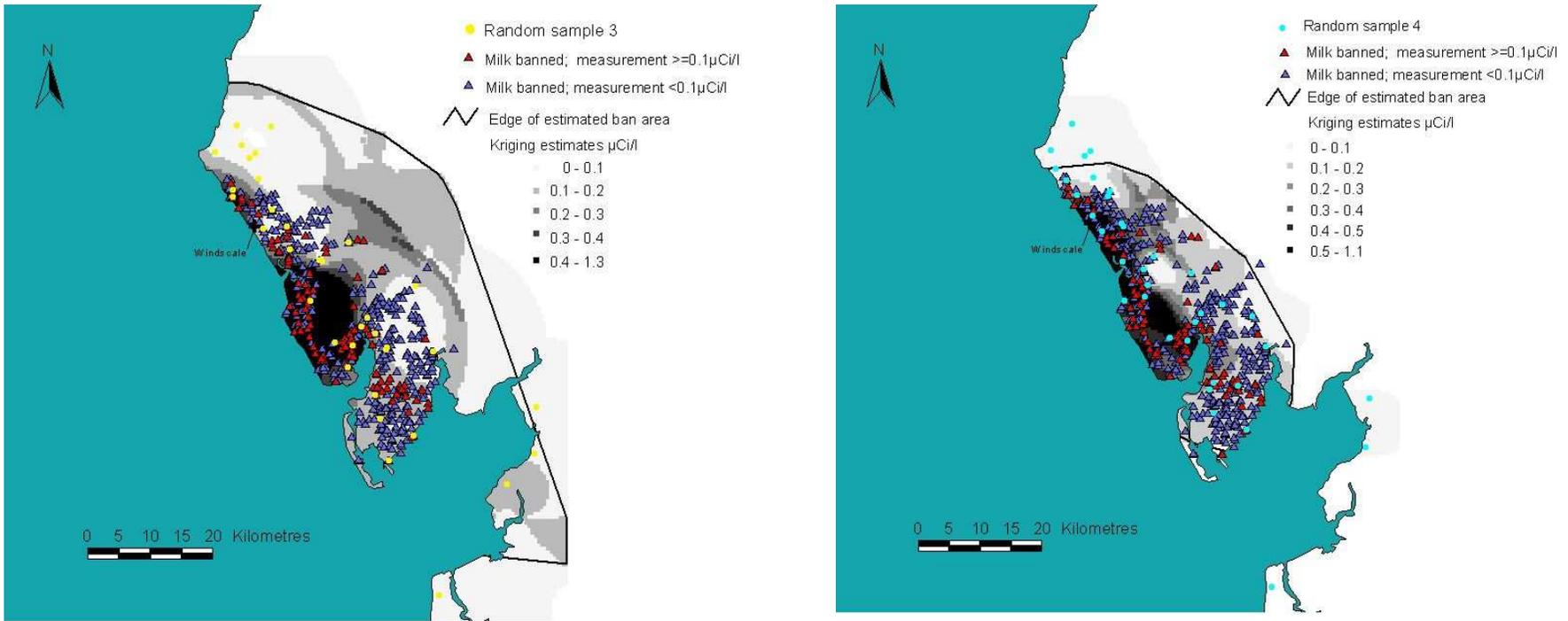
FIGURE 24b Comparison of areas predicted to exceed the milk ban criterion used in 1957, 0.1µCi l^{-1} (3700Bq l^{-1}), with the location of restricted farms derived using random samples 3 and 4 of four randomly located selections of milk measurements.

To give an indication of the errors in the ordinary kriging estimates for Windscale, each subset was also jack-knifed against the remaining locations in the global dataset. The RMS errors for each sample together with a scatter plot of the estimated versus measured values in the jack-knife comparison are shown in Figure 25. The most striking feature of these results is the extent to which ordinary kriging has either overestimated or underestimated the milk activity concentrations. The absolute RMS errors are all large (compared with the mean of the global dataset, 0.195 μCi l^{-1} (7215 Bq l^{-1})) and the spread of values shown by the scatter plot indicates the seriousness of the errors. If comparisons are made with the errors obtained for Tomsk and Bryansk, it appears that ordinary kriging has been much less successful at achieving good estimates with the Windscale data. It might be expected that ordinary kriging would not fair as well with Windscale data as it does with Bryansk data due to the greater influence of a dispersion derived trend. However, there are a number of other important differences between this data set and the ones studied in previous sections. Firstly, only 36 locations[*] are available (at most, assuming that they can all be kriged) at which to calculate the errors. This means that one or two extreme error values can have a strong influence on the RMS error for the sample. Figure 25 also shows the errors when the three locations with the largest individual errors are excluded. Secondly, the Windscale data are less reliable in the sense that there may be factors influencing the results, which cannot be quantified due to a lack of information. For example, it is not known whether any of the cows were fed on stored (uncontaminated) feed, rather than grazing pasture. These factors could have contributed to the observed situation in the global dataset where a location with a very high activity concentration can be found very close to one where the activity concentration is very low. While this does not happen often enough to have a particularly detrimental effect on the variograms, it can strongly influence the kriging estimates. Thus, it is likely that if a high (low) value point is in the data subset and a nearby low (high) value point is not, the jack-knifing algorithm will return an overestimate (underestimate) for the low (high) value point. This is simply because the nearby high (low) value has been given a large weighting. As the graphs illustrate, this only needs to occur at a few locations to produce an RMS error for jack-knifing which is extremely large. The reliability of ordinary kriging is discussed again in Section 6.2 using a similar testing scheme to that employed for Tomsk and Bryansk.

For the purposes of determining food restriction areas, the errors at individual locations are less important than the overall area estimated. These experiments have demonstrated that ordinary kriging can be useful in establishing an initial restriction area that can then be modified as more data becomes available.

---

[*] These are the locations which are in the global data set of 73 locations but are not in the sample of 37 locations. At the sample locations, kriging reproduces the measured values exactly. Errors can only be calculated for locations where the kriging algorithm is interpolating between measured values in the sample.
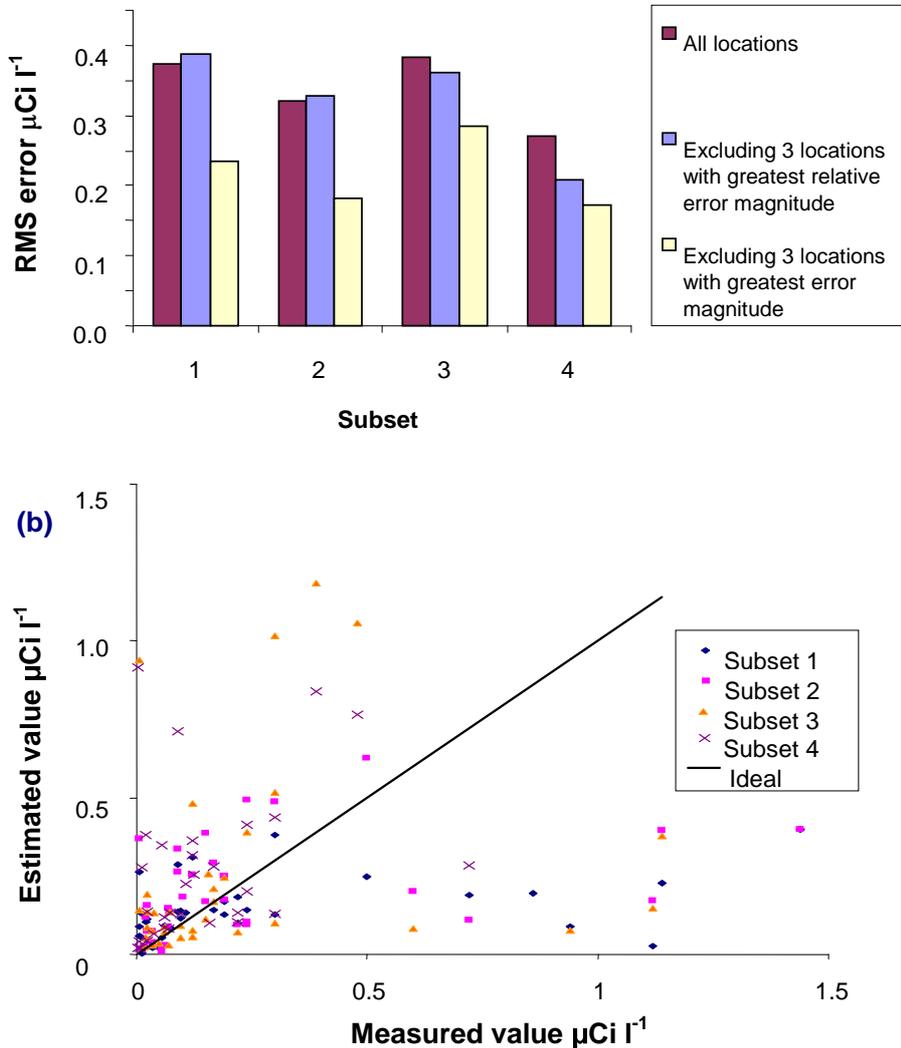
FIGURE 25 (a) Absolute jack-knifing errors for ordinary kriging estimates of activity concentration in Windscale milk and (b) scatter plot of the estimate versus measured jack-knife data for each of the subsets used.

6.2 Reliability of ordinary kriging

In order to compare the application of ordinary kriging to the Windscale fire data with the trials conducted for Tomsk and Bryansk (discussed in Sections 4.1 and 5.1 respectively) a similar set of trials were undertaken. In this case, a series of jack-knife tests of ordinary kriging using three different sample sizes, with three samples (a, b and c) of each sample size, were conducted. A data set of 228 milk measurements taken on the 22nd October was used as the reference global data. This is a smaller number than was available for either Bryansk or Tomsk but is still substantial. However, unlike those examples the measurements are not directly related to ground contamination but to

mobile samplers of contamination ie cows. As discussed in Section 6.1, there is no information on the amount cows were moved around in the days following the accident, or on whether any farmers were able to supply their animals with uncontaminated feed.

The mechanics of the testing were similar to previous trials with random sub-samples of 20, 40 and 60 points taken from the global data set and the remaining data then used as the jack-knife comparison data set for the sample. The root mean square of the difference between the estimate produced by the kriging algorithm and the measured value was calculated, and the results are given in Table 1.

TABLE 1 RMS Error of Kriging Estimates nCi l^{-1}

Sample Size	a	b	c
20	58.51	56.51	47.75
40	60.35	48.10	45.66
60	48.86	49.20	45.49

To relate the results found for Windscale to the comparison between Tomsk and Bryansk shown in Figure 20, the relative RMS error was calculated and is shown in Figure 26.
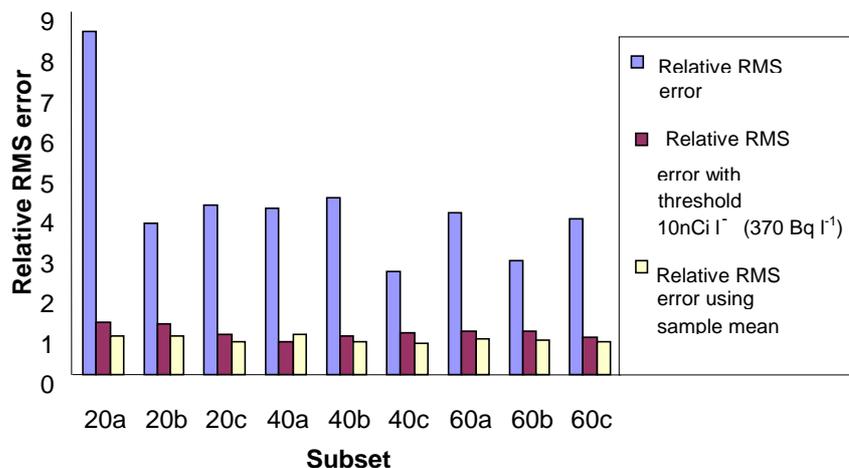
FIGURE 26 A comparison of the error between measured and predicted values for a range of random samples illustrated using three different error measures. The relative RMS error calculation divides each difference between sample and estimate by the sample value or overall mean, as indicated, before performing the RMS calculation.

The relative error found for each of the Windscale samples is very much greater than found in the previous studies. However, as Figure 26 indicates, the major contribution to these large errors is the overestimation of a few very small values. This could arise through a number of mechanisms, as there is ignorance about where cows grazed with respect to where they were measured and if any farmers supplemented or substituted for the contaminated grass their cows would otherwise eat. If the low valued outliers are

neglected, the quality of the estimation is consistent with previous findings, the error is slightly larger than found for Bryansk, but smaller than for the highly trended Tomsk data.

Table 2 gives an indication of the strong dependence of the areal coverage on the configuration of the sub-samples. The better the distribution of sample locations the more successful the kriging algorithm is likely to be at predicting values within the interior of the sampling region and the larger the region that can be estimated. This is a consequence of minimising the collection of redundant information by avoiding taking samples that are either too close to existing samples and contribute little that is new or too far away and isolated from other measurements. Table 2 shows the proportion of potential jack-knife points available from the global data at which the kriging algorithm was able to generate an estimate.

TABLE 2 Proportion (%) of Estimated Points from Kriging Algorithm

Sample Size	a	b	C
20	60	93	68
40	93	90	84
60	93	95	92

As might be expected, the coverage obtained from samples of only 20 points was usually smaller and the RMS error greater than for the other two sample sizes. However, the samples of 60 points did not significantly outperform the samples of 40 points on either of these measures. Thus, for this study, 40 sample locations are sufficient for ordinary kriging to produce accurate and useful results.

6.3 Simulations

The four random samples of 37 milk measurements selected for ordinary kriging in Section 6.1 were also used in a series of simulation calculations. As with the Bryansk data, the sequential Gaussian simulation routine supplied in GSLIB (Deutsch and Journel, 1998) was used. Variograms were produced for the normal-score transformed samples using the same direction for the principal anisotropy axis (125°) as the variograms used for ordinary kriging. These empirical variograms were modelled using a spherical model and a nugget, with the additional constraint that, for normal-score transformed data; the sum of the nugget and sill of the variogram must equal one.

The parameters obtained from these variograms were used as input to the simulation algorithm, and for each sample, 100 realisations were generated. For each point on a 0.5km grid covering the required estimation area around the site, the mean, relative

variance[*], 95th percentile value[†] and the number of realisations whose value exceeded the milk ban criterion were calculated.

The individual results for one of the samples are shown in Figure 27[‡]. It can be seen from the map of the variances, Figure 27(b) that, as expected, the variance is highest in those regions that are remote from measurements ie the region beyond the convex hull that can be envisaged from the sample locations. Each of the 100 realisations computed using this particular sample will have similar values near the sample data. At locations further away, where the sample data have less influence, the values of any individual realisation are increasingly likely to differ from those of other realisations and the variance is therefore greater. Simulation can be used to indicate areas where the maximum additional information can be gained from a fixed number of further measurements. The areas with the highest variance are those where there is least knowledge about the true data values.

The map of 95th percentiles, Figure 27(c), gives an indication of whether the measured value at any location will exceed the ban criterion. In the green areas where the 95th percentile estimate is between 0 and 0.1 μCi l^{-1}, the probability of exceeding the criterion is no higher than 5%. It can be seen that these areas contain several sample locations, each of which is well below 0.1 μCi l^{-1}, so it is possible to be confident that the measured value at a nearby location will be similarly small. As the value represented by the 95th percentile increases, the probability of exceeding the ban limit also increases. It will be noted that the map shows this probability increasing with distance from the site in a North-Easterly direction, a result which appears unrealistic. It may be expected that at the distances from the release point shown, the opposite effect should be observed. This effect is an artefact stemming from the lack of measurement data at this distance; since it has no real data available the simulation algorithm returns a value for these locations which is close to the global mean value.

The statistics produced by the simulation become more reliable as an increasing number of realisations are used to generate them, although the rate of improvement diminishes for very large numbers of realisations. Due to time constraints, it is useful to determine a cut-off point beyond which it is deemed inefficient to continue producing more realisations. A basic test to assess the reliability of the 100-realisations approach is to compare the graphical output from the simulation based on both 100 and 1000 realisations. The maps of the means are shown in Figure 28. They illustrate that, for practical purposes, the difference between the results for the two simulation runs is negligible. As one might expect, the map based on 1000 realisations is smoother, but it may be expected that a decision-maker would come to similar conclusions irrespective of which map is used.

---

[*] variance divided by square of the mean
[†] 5th highest value from the 100 realisations
[‡] Results from the other three sub-samples are broadly similar to those illustrated in Figure 27
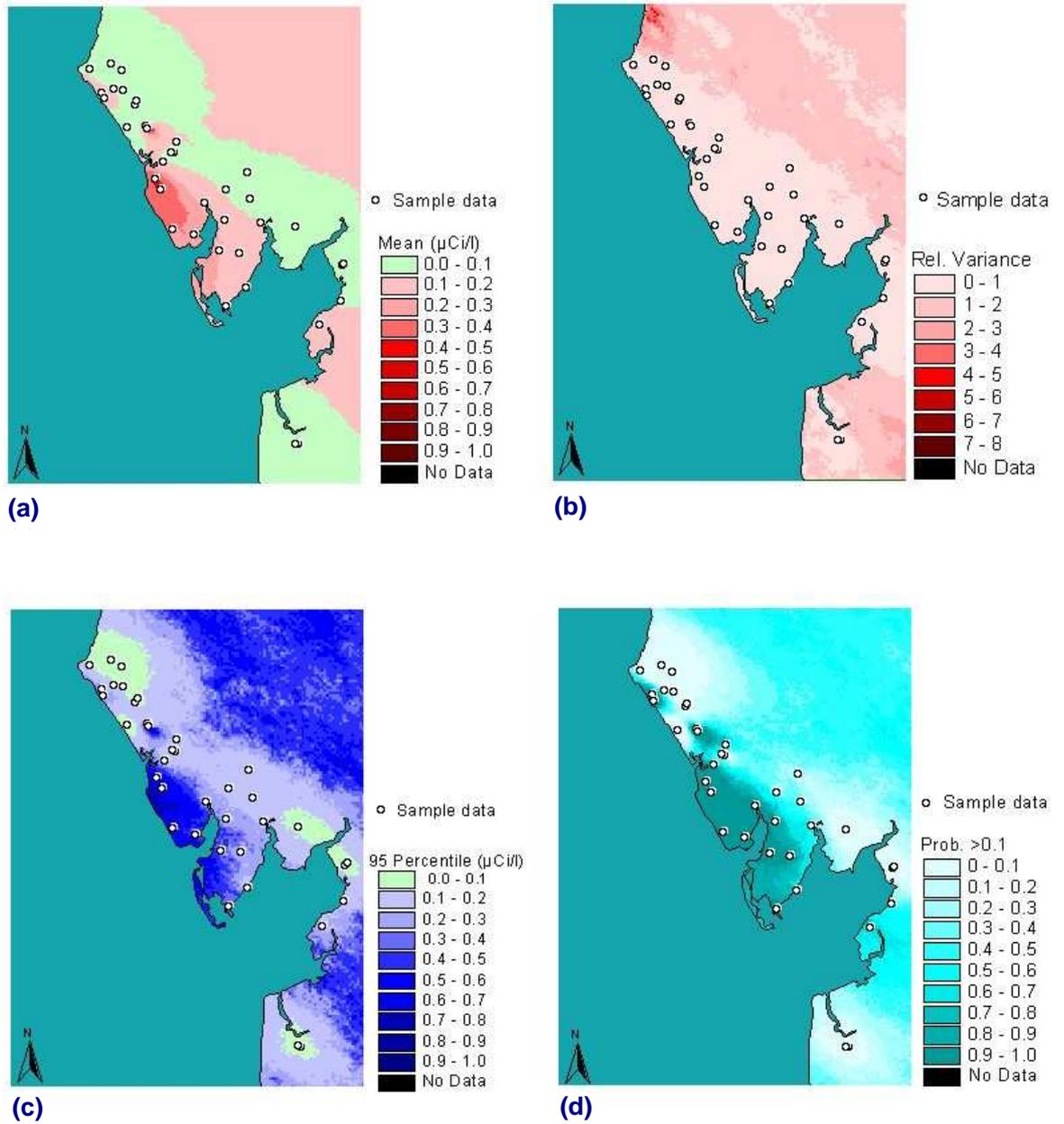
FIGURE 27 Maps generated from 100 realisations to show (a) the mean estimate, (b) relative variance, (c) 95th percentile, and (d) probability of exceeding the Windscale milk ban criterion of 0.1 μCi l^{-1} (3700 Bq l^{-1}).

(a) (b)
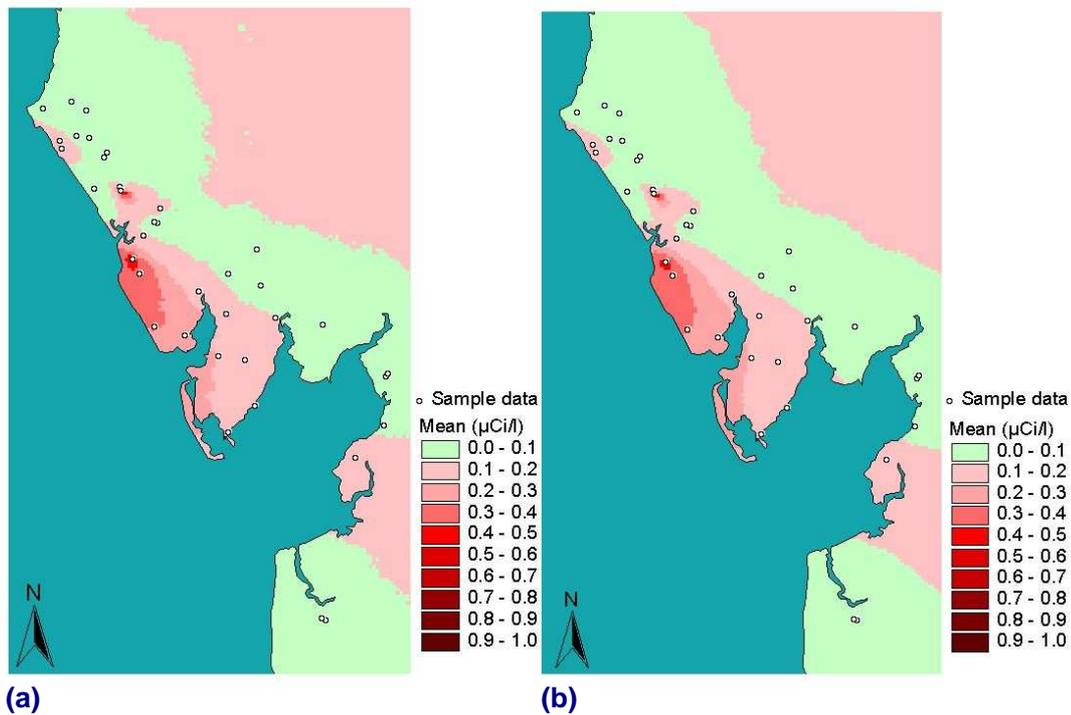
FIGURE 28 Comparison of mean activity concentration in milk estimated using (a)100 and (b)1000 realisations.

A more extensive discussion of the effect of using different numbers of realisations in a simulation is given in Appendix C.

6.4   Reliability of Gaussian simulation

To assess the reliability of the sequential Gaussian simulation calculations the procedure discussed in Section 3.1.4.2 was applied to the sub-samples introduced in Section 6.2. The jack-knife data are used to generate accuracy plots for each of the subsets. The results shown in Figure 29 indicate a general improvement in the simulations as more sample data are used, with the curves becoming progressively closer to the reference ideal as the sample size increases.

To provide a simpler numerical assessment of the Gaussian simulation output, a measure of the uncertainty in the distributions was calculated by finding the average variance of the output distributions at the jack-knife data points. The results are given in Table 3.

TABLE 3 Uncertainty of Simulation Distributions

Sample Size	a	b	C
20	$4.32 \ 10^{-3}$	$5.58 \ 10^{-3}$	$2.07 \ 10^{-3}$
40	$1.66 \ 10^{-3}$	$2.71 \ 10^{-3}$	$2.24 \ 10^{-3}$
60	$5.29 \ 10^{-3}$	$1.35 \ 10^{-3}$	$1.49 \ 10^{-3}$
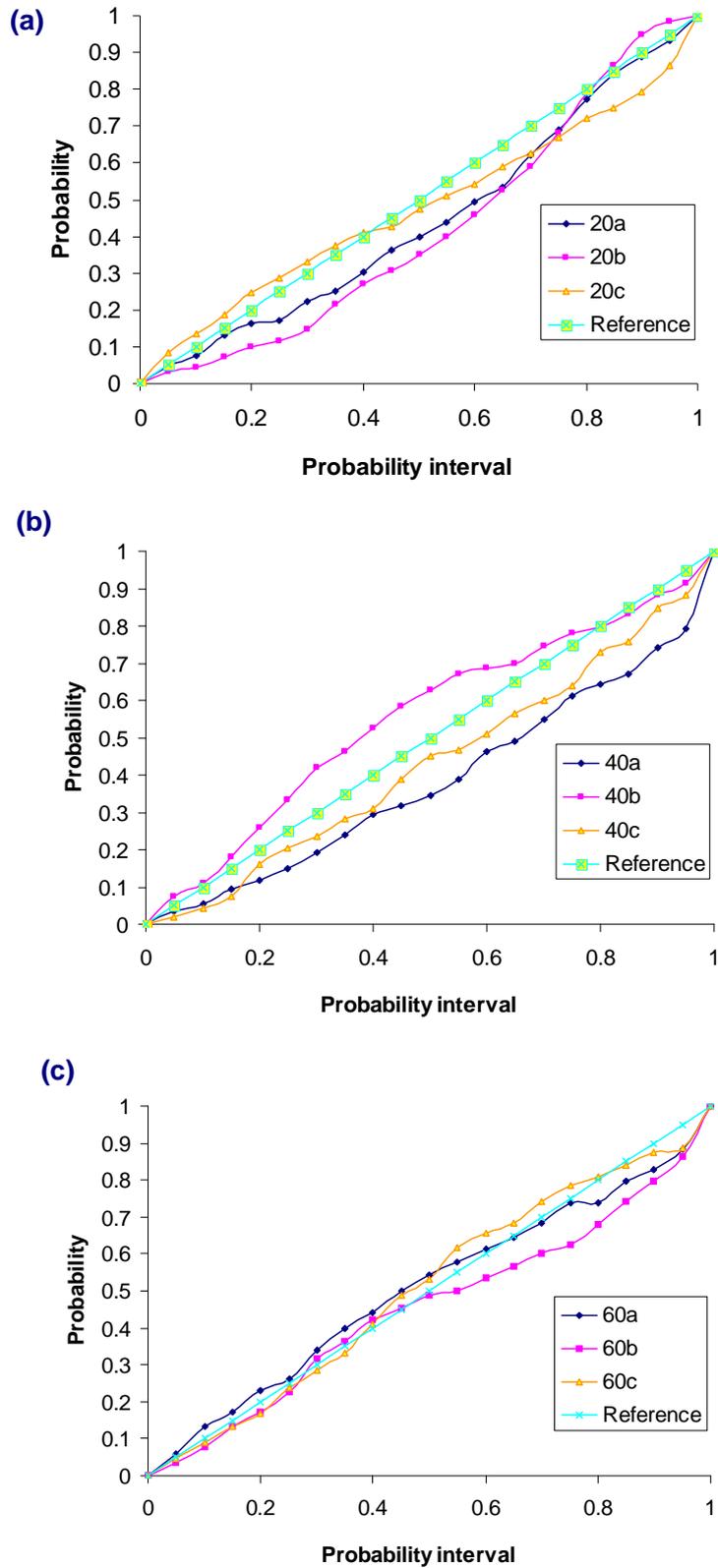
FIGURE 29 Accuracy plots for three subsets each of (a) 20, (b) 40, and (c) 60 data. The closer the results are to the reference lines shown the more accurate and precise the simulation. Results above the reference lines are accurate but not precise and those below are not accurate.

The average variance shown in Table 3 is sensitive to outliers in the data but the results nevertheless indicate that apart from the anomalous sample 60a, (see also Figure 26) there is an overall decrease in the uncertainty of the simulated distributions as more data points are used. The more substantial results shown in the plots in Figure 29 indicate that the samples of 60 points have greater accuracy and precision than the other two sample sizes. There is however, no significant difference between the accuracy and precision of the samples of 20 and 40 points. For this study, it could be concluded that at least 60 sample locations are necessary to produce accurate and precise output, and that the usefulness of the output improves as more locations are used.

6.5   Cokriging

The variety of sample types in the Windscale database provided the opportunity to test another geostatistical technique: cokriging. This allows secondary data, which are correlated with the primary variable of interest, to be included in the kriging algorithm. This, as discussed in Section 3.1.3 and in more detail in Appendix B, is a potentially useful approach if the variable of interest is undersampled. It was decided to retain the ^{131}I in milk measurements as the primary data, given that they would be quantities of great interest to the Agency when considering food restrictions. In addition, previous work has established the reliability of these data. As cokriging estimates will only be an improvement over those of ordinary kriging when the primary variable is appreciably undersampled the principal consideration when selecting a secondary variable is the number of measurements available. On this consideration alone, the gamma dose rate measurements appeared to be the most promising candidate. However, there are several difficulties associated with using the Windscale gamma dose rates, the most important of which stems from the use of the 1413A meter to measure the great majority of them. This device has a highly non-linear energy response, as shown in Figure 30. This problem is compounded by ignorance of the precise combination of radionuclides contributing to the dose rates. In addition, the correlation between ^{131}I in milk and the gamma dose rates could vary spatially as a result of the differing deposition velocities of iodine and the aggregation of other radionuclides represented by the gamma dose rate measurement. Use of non-contemporaneous milk and gamma measurements would add to the complexity through the introduction of a time-dependence to the existing spatial dependence of any correlation, arising from the differing decay rates of the various radionuclides.

It is possible that the Windscale gamma measurements could be partially salvaged. However, this may require additional data and would only provide, if successful, very uncertain estimates of dose rates for use in co-kriging calculations. In addition the production of the dose rate estimates would probably require a lot of computational effort. An alternative secondary variable for the SECTAR investigations was therefore sought. The measurements of iodine on grass, although much less numerous than the gamma dose rate measurements, were therefore selected. It should be noted that the difficulties inherent in interpreting and using the Windscale gamma dose rate estimates are essentially historical. Improved instrumentation and the greater availability of

gamma spectroscopy to provide a breakdown of the radionuclides contributing to the measured dose rate at particular locations would remove most of the uncertainty.
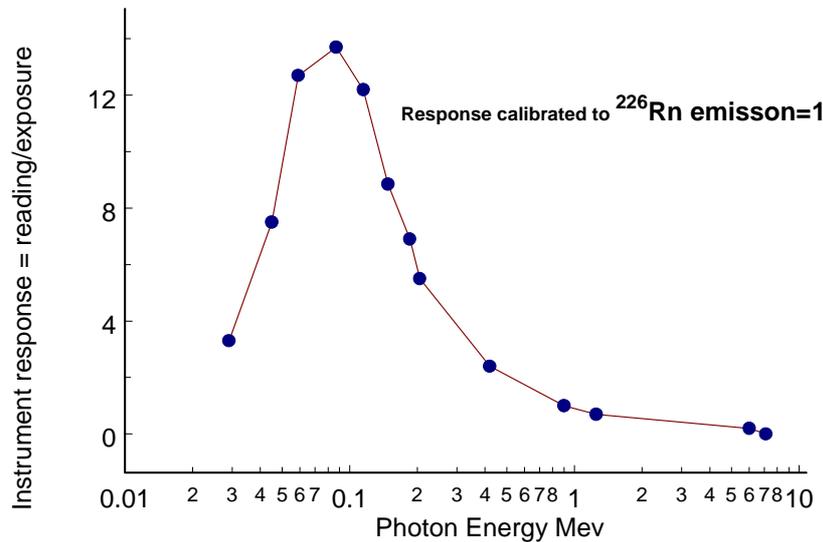
FIGURE 30 Energy response of 1413A survey monitor.

Prior to specifying the exact data sets for the experiments, the possibility of meeting a secondary objective was considered. This would help to demonstrate the potential of co-kriging generally and its application to a particular problem ie providing an estimate of the extent of the required Windscale milk ban at an early stage. Direct comparison could then be made with the Windscale ordinary kriging experiments. Ideally, co-kriging would allow the Windscale milk ban to be derived using the limited number of iodine in milk measurements available on the day after the end of the release (13th October). These data would be used in conjunction with measurements of iodine on grass made on the same day. The use of the grass measurements as a secondary variable would avoid the need to use data from later dates required in the ordinary kriging trials (see Section 6.1).

Unfortunately, too few deposition measurements were made in the early stages of the accident for this calculation to be performed. The alternative approach is therefore to concentrate on having a large number of milk measurements as a global data set to allow the success of the co-kriging trial to be assessed. Thus, a random sample of 40 measurements of ^{131}I in milk was selected to be the primary data in the co-kriging trials. These were drawn from a global data set of 228 distinct locations where milk was sampled 10 days after the end of the release on 22nd October. These global data were also used in Section 6.2 in the standard tests of ordinary kriging applied to all the available accident data. A small amount of secondary deposition data had accumulated by this stage with deposition measurements available from 82 locations. However, 13 of these were at least 150 km from the site and were excluded from further consideration,

as they were simply too far away to influence kriging estimates near the site. Thus, the grass data available as secondary support consisted of the remaining 69 measurements decay-corrected[*] to the 22nd October.

The variogram surface for the milk data showed clear anisotropy, with the principal axis at an angle of approximately 130°, which is consistent with the 125° principal axis previously determined for the ordinary kriging data of Section 6.1. Given that the same underlying processes of atmospheric dispersion and deposition are behind both grass and milk measurements, it was expected that the deposition data would exhibit similar anisotropy. However, the variogram surface found for the deposition measurements was less conclusive, with the observed anisotropy strongly influenced by an anomalously high measurement about 6km south-east of the site. Exclusion of this measurement resulted in a better behaved empirical variogram but with almost isotropic behaviour, whereas retaining it produced a variogram with more scatter, but where the anisotropy was similar to that found using the milk data. As there was no known reason to exclude this measurement, it was retained. However, this made the fitting of a model variogram more difficult. Fitting was expected to become easier if the grass data were block kriged to give them a level of support similar to the milk measurements.

Calculation of mean and variance for the two sample types showed that the relative variance of the grass data was much larger. This is to be expected as the milk measurements are effectively averaging the deposition over a large area, as they are related to the deposition over the area grazed by a herd of cows. This spatial averaging should have a smoothing effect and would account for the reduced variance. The grass measurements are likely to have been taken over a small area, of the order of one square metre, and so their level of support is quite different from that of the milk data. To relate the two sample types, it was considered sensible for them to be on a similar level of support, this was done using block kriging (see Section 3.1.2.4).

The block size chosen was a square of side 0.5km, as it was considered that this was a reasonable estimate for the area which could be grazed by a herd of cows. A map of the area shows that fields tend to be somewhat smaller than this but there is no reason to assume that the herd would be confined to a single field over several days. Square blocks were chosen to eliminate any 'artificial' anisotropy effects, which may arise if the data were smoothed more in one direction than in another. The sides of the blocks were aligned parallel to the assumed anisotropy axes, again to avoid distortion of the natural anisotropy, but it is now thought likely that any distortion caused by not arranging the blocks in this way would be minimal.

Examination of the block kriged estimates showed the expected reduction in relative variance. Each of the 69 grass point measurements was replaced with the estimate calculated over the appropriate block; resulting in estimates for all but two of these measurements[†]. In addition to providing a common level of support, block kriging also

---

[*] Note that no corrections were made for other time-dependent effects eg migration down the soil column. The half life was taken to be 8 days.
[†] At least 2 measurements within the search radius were required to krig a block, so not all blocks could be estimated.

enabled the creation of a number of deposition estimates that were collocated with the primary milk measurements. These are required to generate a cross-variogram (see Section B3), and as might be expected Windscale measurements were not naturally collocated. A decision was taken that a grass block estimate and a milk measurement would be considered collocated if the milk measurement was contained within the extent of the grass block area. Any deposition block that contained the location of a milk measurement from the random sample was therefore added to the data used to model the cross-variogram. This provided 19 blocks where the deposition value at the milk measurement location was taken to be the block estimate for the block that contained it.

A summary of the final data selection for cokriging is given in Table 4.

TABLE 4 Data for cokriging

Description	Number of data points
Primary variable: random sample from measurements of iodine-131 in milk made on 22 October	40
Secondary variable: block kriged estimates of deposition at locations of original grass measurements or locations in the milk random sample	86
Cross-variogram data: collocated milk random sample and grass block kriged estimates	19
Cokriging input data: all primary and secondary data	107 (40+86-19)

Directional variograms and cross-variograms with a principal anisotropy axis of 130° were created using VarioWin and modelled to satisfy the linear model of coregionalisation (see Section 3.1.3). A nugget and two spherical structures were used for each model, one structure having a range of 8 km and the other, 32 km. A trial run using traditional ordinary cokriging, with two nonbias conditions, resulted in many locations being left unestimated, as there were no primary data within the search radius. It was therefore decided that standardised ordinary cokriging was more suitable, since its single nonbias condition can still be satisfied when only secondary data are available. To allow rescaling of the data according to Equation B4 in Section B3, milk and grass data were converted to common units of μCi kg^{-1}. This required the assumption of a milk density of 1 kg l^{-1} and pasture herbage density of 0.5 kg m^{-2} (Simmonds, 1985). The primary and secondary means were taken as the arithmetic means of the sample data.

Figure 31 (a) shows that the measurements are separated into two quite distinct areas; the 'main area' around the site containing primary and secondary data, and an 'arc' about 100 km south-south-east of the site which only contains secondary data. Although standardised ordinary cokriging estimates could be calculated for each area, estimates using only the 'arc' data will not be reliable. This is because there is a lack of primary data in this area with which to calculate a suitable mean value for use in Equation B4. The local mean for the primary data would therefore have to be estimated, for example by assuming that the ratio of the primary means in the two areas was the same as the ratio of the secondary means (which can be calculated from the data). For

this reason, cokriging estimates have only been calculated for the 'main' area where primary and secondary data are available.

Estimates calculated by cokriging over the main area were compared with the results obtained through ordinary kriging using the milk random sample alone, shown in Figure 31(b). A simple comparison by eye reveals that there is little difference between the results. However, co-kriging had the effect of simplifying the boundaries between areas of different milk concentration by providing the data necessary to remove gaps in the coverage. The kriging errors were evaluated at locations where milk measurements had been made on 22nd October, excluding those locations in the random sample itself. (By definition kriging is an exact interpolator and the errors at the random sample locations will be zero). Note that this is essentially the same as the jack-knifing procedure used with the Bryansk and Tomsk data. Calculation of the RMS error for each of the kriging techniques confirmed that cokriging offered little benefit in this case, with both methods having the same RMS error ie 0.048 μCi kg^{-1} (1776 Bq l^{-1}). However, cokriging did provide an estimate at locations where ordinary kriging lacked sufficient data to do so, and the mean absolute cokriging variance at the jack-knifed locations was smaller than the mean absolute ordinary kriging variance, 0.002 μCi kg^{-1} (74 Bq kg^{-1}) as opposed to 0.071 μCi kg^{-1} (2627 Bq kg^{-1}). The overall result would suggest, however, that there were insufficient secondary data in this experiment to take advantage of the improvements possible using cokriging.

A complication that has not been considered in the application of co-kriging is the role time might play in determining the effective correlation between different data types (see Section 7.2.3). For example, the concentration expected in milk might be more strongly related to the activity on the grass over a period of a few days, a few days prior to the milking taking place. This is unlikely to make a dramatic difference in this case, as the activity in the grass is likely to be simply related to the activity present a few days previously. However, if it had rained in a part of the area of interest a few days before it may be appropriate to subdivide the problem into areas where different correlations are expected. An alternative and potentially simpler approach to the use of supporting data is discussed in the Section 6.6 which allows for time including times in the future to be considered.
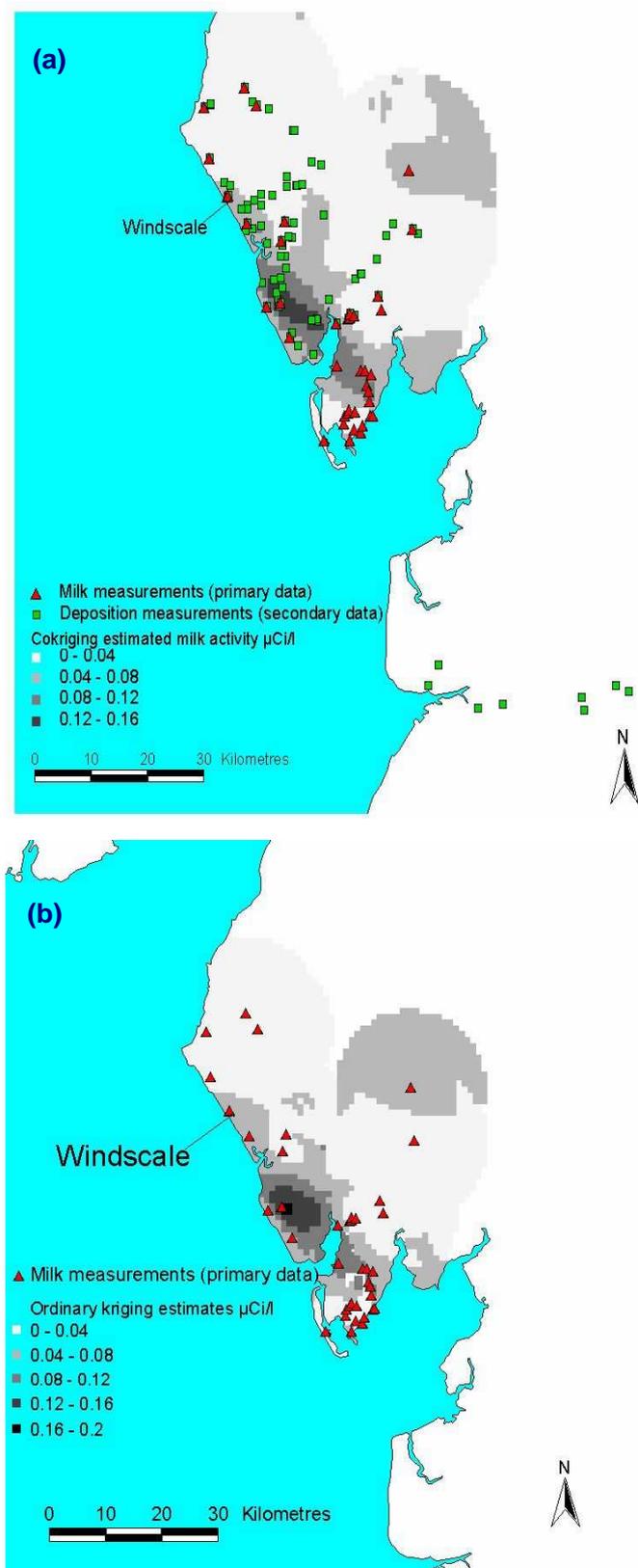
FIGURE 31 Locations of data and areas estimated by (a) cokriging and (b) ordinary kriging.

6.6   Geostatistics and foodchain modelling

Cokriging requires complex and, without convenient software, time consuming procedures to be undertaken to produce an estimate. In many cases, a simpler method of producing an estimate of the concentration in one material using information about another would be to use existing models representing the transfer of radioactivity through the environment. For example, if measurements of deposited activity on grass are available, estimates of the resulting activity concentration in milk can be calculated as a function of time by an appropriate process model. The FARMLAND model (Brown and Simmonds, 1995) used at HPA-RPD is one such model, which represents the transfers between different parts of the environment using a series of coupled first order differential equations with transfer coefficients determined experimentally. The drawback of this approach for emergency assessments is that few measurements are available in the early stages of an accident. This limits the number of estimates that can be made using a foodchain model alone. In addition, any estimates are restricted to the locations of the original measurements. This section discusses work that aimed to combine the spatial estimation techniques of geostatistics with the convenience of a foodchain model to obtain a large number of milk contamination estimates from a limited number of deposition measurements. The errors on the estimates using this method were compared with those from the ordinary kriging of milk measurements.

The first stage in the calculation is to run the FARMLAND model assuming that cows are on pasture, for unit deposition (1 Bq km^{-2}) of ^{131}I onto grass and soil. This allows the effective half-life of the iodine on grass to be determined. However, the Windscale accident was an extended release and it is not clear from the information available exactly when it started and stopped. It is known that the release was not uniform, but the exact behaviour of the iodine is not known (ie whether it followed the same pattern as the general release or came off more rapidly). For simplicity, these trials assume that the release of iodine was continuous between midnight and 1100 hours on 11th October. Thus, all deposition was assumed to have occurred by the end of this period. A more complicated representation, using two peak releases, was also tried but the difference was negligible. Deposition was partitioned between grass and topsoil using an interception factor of 0.25 (Simmonds, 1985) and FARMLAND was run to obtain the time evolution of the activity on grass over a period of nearly 2 weeks after the deposition occurred. This is illustrated in Figure 32, from which the effective half-life of ^{131}I on grass can be estimated to be approximately 3.6 days. This effective half-life was used to decay correct back to 11th October the 69 deposition measurements made on or before 22nd October. These corrected measurements were then block kriged to obtain deposition estimates (for 11th October) at as many as possible of the locations where milk was sampled on 22nd October.
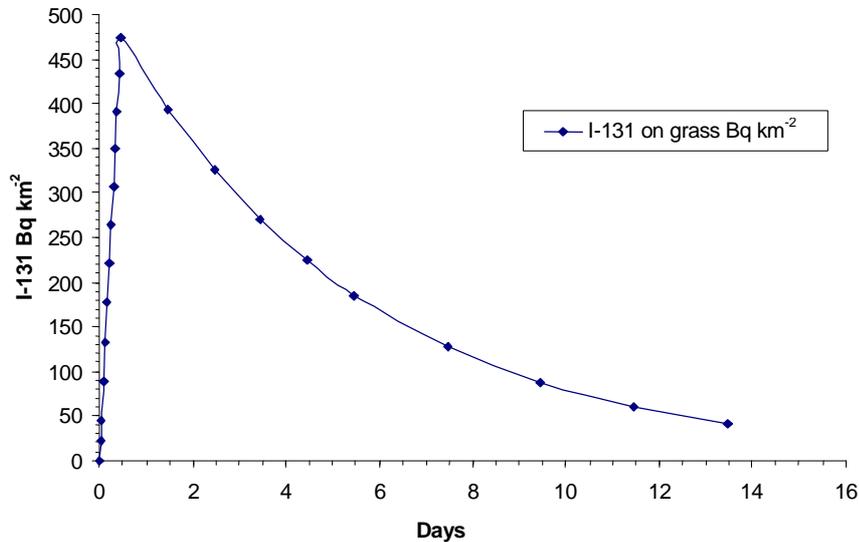
FIGURE 32 FARMLAND model prediction for decay of ^{131}I on grass.

This approach not only extends the number of matching locations but also assumes, reasonably, that it should be better to compare appropriately scaled area averaged grass and soil concentrations with the radioactivity found in milk for the same area. However, this procedure may mislead for several reasons. It assumes that the same half-life applies everywhere. The modelling neglects the effect of rain washing iodine down the soil column and shortening the effective half-life. Rain will further complicate the results by having a different effect depending on when the deposition measurement was made with respect to the rainfall and the time of the milk estimate. There is also some ambiguity in the data as to whether grass measurements also included some soil. Finally, it assumes that the cows have been in the same area for several days. However, the result should be broadly indicative of the activity concentration to be expected in milk. Thus, using the FARMLAND curve of Figure 33, which gives the time evolution of activity concentration in milk for the unit deposit discussed above, it was possible to calculate the activity concentrations on 22nd October corresponding to the block kriged deposition estimates for 11th October. To gain an idea of the effectiveness of this method a comparison with measured results can be made. However, it is also of interest to compare this approach involving two sets of estimations, the use of block kriging and a foodchain model, with results obtained by the ordinary kriging of milk measurements from 22nd October. A jack-knife calculation on the original milk measurements was therefore performed. A random sample of 60 milk measurements were selected from the 129 locations where FARMLAND estimates had been possible (ie. where a grass block estimate had been made) and measured values were available. A jack-knifing calculation using ordinary kriging then gave estimates for milk activity concentration at 49 of the remaining 69 locations (the remaining 20 locations could not be estimated). The RMS estimation error was calculated for both methods at these locations, and the foodchain model approach was found to produce an RMS error

an order of magnitude greater than ordinary kriging of the milk data. (The errors were $2 \ 10^4$ Bq l^{-1} and $2 \ 10^3$ Bq l^{-1} respectively; the milk measurements at the jack-knifing locations had a mean of 2.1 10^3 Bq l^{-1} and a standard deviation of 2.7 10^3 Bq l^{-1}). It should be borne in mind that kriging is an exact interpolator and therefore kriging the milk results is likely to produce a better result simply because the smoothed interpolated surface is constrained to agree with the known values whereas the transformed grass results are not constrained in this way. Thus, at least in this case, it appears that the additional data processing involved in using a more complicated approach is not justified. The estimates provided by ordinary kriging (and by implication the more complex cokriging approach discussed in Section 6.5) are as good (or better) than those from the combined use of kriging and process models. Unlike, for example, cokriging where the supporting data are event specific, each of the additional layers of estimation in the model supported calculation are more likely to contribute additional uncertainty to the final result. However, this approach does have two advantages: firstly, it allows one type of measurement to be transformed into another, not just at the sampling locations but everywhere a kriged estimate can be produced, and secondly, because it uses generic models this is potentially a quick and simple procedure.
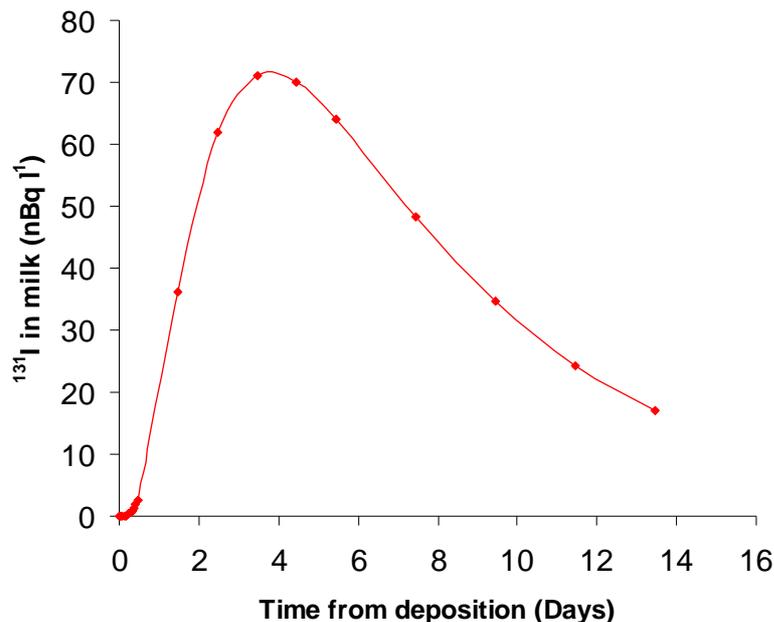
FIGURE 33 FARMLAND modelling of time evolution of ^{131}I activity concentration in milk, following deposition of 1Bq km^{-2} to grass and soil.

6.7   Bayesian analysis of deposition data

The Bayesian atmospheric dispersion program described in Section 3.2 was tested using ^{131}I deposition data. 81 deposition measurements from the Windscale database

were used as the global dataset and the predictions of the model checked by cross-validation with subsets of 20, 30, 40 and 50 points. The cross-validation estimates were then compared with the measured values and the RMS error for each subset was calculated. The error in this case is defined as the difference between the measured value and the mean of the posterior predictive distribution at the measurement location, where the Bayesian method calculates the distribution for the ith point, using all measurements except that of the ith point.

Three prediction strategies were compared, the first relied entirely on the Gaussian dispersion model with fixed values obtained from the literature for the calibration parameters, the second introduced the model inadequacy corrections to this procedure, and the third included Bayesian calibration to give a fully Bayesian analysis. In the first two experiments where the calibration parameters were fixed, the values for deposition velocity and source term for ^{131}I were taken as 4×10^{-3} ms^{-1} and 1000 TBq respectively, from Chamberlain (1959, 1981). Large prior variances (5.0) were assumed in the two Bayesian analyses so that the source term and deposition information, which obviously would not be available in the early stages of an accident, would not have an undue influence on the results. For the full Bayesian method, quantile-quantile plots for the standardised errors showed that they were approximately N(0,1) distributed[*] for all four subset sizes, thus supporting the validity of the model.

The RMS errors for the three methods are shown in Table 5 (from Kennedy *et al*, 2002). It can be seen that the estimates are improved when model inadequacy is included, and improved further still when the full Bayesian calibration is used.

TABLE 5 RMS errors of prediction for Windscale samples

Sample size	Gaussian dispersion model only	Fixed calibration parameters but including model inadequacy	Model inadequacy and Bayesian calibration
n=20	2.98	1.93	0.94
n=30	3.09	1.79	0.87
n=40	2.85	1.80	0.96
n=50	2.83	1.95	0.88

The reason for the improvements shown in Table 5 as more Bayesian attributes are used is that a simple Gaussian plume model is a poor representation of the actual dispersion process during the Windscale fire. It does not fit the observed data well. Running the model for the 50-point dataset using only the Bayesian calibration and not the inadequacy term showed that the Gaussian plume model had minimal influence on the predictions. The observed results were almost entirely generated by the action of the model inadequacy component. This explains why the errors found for the second experiment (model inadequacy with fixed calibration parameters) are smaller than those found for the first ie. the trial of an unaided Gaussian dispersion model. The further reduction in errors observed in the third experiment, when Bayesian calibration is

---

[*] normally distributed with a mean of zero and a unit variance.

included, is accounted for by the calibration process modifying (reducing) the influence of the badly fitting Gaussian dispersion model on the predictions and thus improving the results.

A disadvantage of the full Bayesian analysis is the potential loss of identity of the source term scaling the Gaussian dispersion model (see Section 3.2.3). In this case, the influence of the dispersion model is automatically reduced if the Bayesian model controlling the overall evaluation detects that the dispersion model is acting to distort the results. If an estimate of the true source term, at least of those radionuclides that deposit, were required from the full model, an integral summing all the estimated deposits would have to be carried out.

The advantages of the Bayesian dispersion model in this case are clear; it allows the measurement data to 'correct' the Gaussian dispersion predictions to obtain better deposition estimates, as more data become available. In areas where there are few measurements, the dispersion model will still allow estimates to be made although this lack of data will be reflected in a larger variance for estimates at such locations. This ability of the model to give an indication of the uncertainty associated with its predictions could be of assistance in deciding the best allocation of monitoring resources to reduce the uncertainty.

7   DISCUSSION AND RECOMMENDATIONS

SECTAR has investigated the application of a number of statistical techniques to the problem of assessing the extent of contamination shortly after a nuclear accident. Statistical techniques require measurement information; thus there will inevitably be some initial period during which there will be insufficient data available for the techniques to be used. However, it was not clear at the beginning of this research programme how much data would be required for the techniques to be used. It was also not clear if the techniques investigated would provide the sought after improvements in the understanding and assessment of accident consequences. SECTAR has demonstrated that insights and predictions can be gleaned from only a few tens of measurements. This is fortunate as there are ultimately only three ways to improve estimates of the extent of contamination: enhance process modelling to more successfully predict levels found in the environment; apply statistical methods to make better use of the available information from both measurements and models, and undertake a more rapid and extensive measurement campaign. The best approach is probably the last but they are all complementary and the least studied in this context is the use of statistical methods. SECTAR has demonstrated that statistical techniques can be used with comparatively small amounts of data to give results that are consistent with the understanding gained later through further measurement. SECTAR has therefore demonstrated that in principle these techniques could be used to help decision-makers decide on the extent and timing of appropriate countermeasures.

The questions that the research addressed were:

a   Could statistical techniques using limited data provide an improvement over existing methods based on simple model estimates?

b   Could statistical techniques be applied sufficiently easily and rapidly to be effective?

c   Could statistical techniques handle the complexity of real accidents?

d   Could they provide information on the uncertainty in estimates and the likelihood of exceeding thresholds?

e   Could statistical techniques be used in conjunction with simple model estimates?

The SECTAR project has attempted to provide answers to the above questions within the limitations imposed by the amount and form of post accident data available to test the techniques. The three sets of accident data obtained, fortunately from the perspective of the project, cover a broad range of possible scenarios:

a   Tomsk: a short duration release well represented by conventional dispersion modelling of the type regularly used in emergency response exercises.

b   Bryansk: a large release with a complex far-field deposition pattern.

c   Windscale: a release lasting several hours in changing meteorological conditions with dispersion occurring over complex terrain.

It should be noted that for each of the accidents investigated it was possible to take a small sample from the global accident data set and to use this sample to produce predictions and estimates that were consistent with the remaining unused measurements (see Sections 4, 5 and 6).

7.1   Performance of SECTAR methods

SECTAR has used data from three accidents to assess the merits of a variety of techniques. The complexity of the real data used, the techniques presented and the necessarily complex arrangements required to assess the techniques can obscure the essential findings. It is therefore appropriate to consider how successful the techniques investigated under the SECTAR project might be at meeting the needs of decision makers by considering the answers to the questions of Section 7 in some detail.

7.1.1   Improvement over simple model estimates

Simple model estimates will always be central to assessing the consequences of an accident in the early stages. In some circumstances eg the accident at Tomsk (see Section 4) they may provide a wholly adequate description of the dispersion and deposition of the contamination that covers the lifecycle of the event, except possibly for detailed studies of the long term environmental and ecological follow-up of the release. Nevertheless, some of the statistical techniques tested against Tomsk data did provide as good or better a representation of the measured reality than the simple dispersion models employed. However, the improvements were small and difficult to judge

conclusively because of the uncertain effect of the partly censored data. Thus, although kriging with a trend was superficially the best performing of the techniques tried when estimating deposition near the centre of the plume path the advantages of applying the technique in this sort of application are not clear-cut.

Unsurprisingly, the Gaussian dispersion model that incorporated the use of Bayesian assimilation generally performed as well as a standard Gaussian dispersion model fitted to the Tomsk data (see Section 4.3). However, by integrating Bayesian assimilation within the dispersion modelling process (see Section 3.3) the model is also capable of providing reasonable results when the plume deviates from the Gaussian dispersion ideal represented by the Tomsk data. For example, the Bayesian technique produced results that showed a greater fidelity to the Windscale measurements than could be achieved using a simple dispersion model. (See the comparisons of a fitted Gaussian dispersion model with a Bayesian analyses in Sections 4.3, 6.7).

Ordinary kriging, although an inappropriate technique for such highly trended data as Tomsk, would be applicable to the assessment of far field contamination as illustrated by the Bryansk results of Section 5. If the UK were to be contaminated following an overseas release, this technique would complement and refine the predictions available from the Nuclear Accident Model (NAME) run by the Meteorological Office (Ryall, 2000) by helping to confirm the situation on the ground using only a few tens of measurements in an affected area. The use of ordinary kriging is also supported by the results of the Windscale trials of Section 6 where the technique was used to estimate the extent of the milk ban required after the Windscale fire, although there was excessive conservatism introduced into the prediction if the measurement locations were not adequately spread over the area.

Co-kriging is a technique that might be of use in particular circumstances ie were there are comparatively few measurements of the data of primary interest available in an area with a large amount of other data that are expected to be correlated with the measurements of interest. Unfortunately, the data available for testing the potential of the technique within the SECTAR project were limited. However, applying the technique to the non-optimal mix of data available produced results that were little better than ordinary kriging (see Section 6.5).

Simulation, although limited to situations where simple or ordinary kriging may be used, will provide information on uncertainty. This complements the uncertainty information available through a Bayesian analysis in circumstances when ordinary kriging is inappropriate. Simulation shares the preliminary calculation requirements of ordinary kriging and therefore acts as a natural extension to the procedure with the discussion of Section 6.4 indicating how the accuracy and precision of simulations are likely to improve as more data become available.

Thus, Bayesian assimilation, ordinary kriging and simulation have been demonstrated to work under the conditions described above and by their nature may be expected to make progressively better predictions as more data become available. All these techniques provide more information than simple models. In the case of Bayesian assimilation, this is because information on the uncertainty of the estimate is always provided with the result, even if the best estimate is similar to that of a simple dispersion

model due to lack of data[*]. Ordinary kriging automatically provides more information when it is used to estimate the pattern contamination from an overseas release as there is no alternative simple model available. More generally, the kriging variance will indicate the regions where estimates are likely to be uncertain and simulations will provide uncertainty estimates showing the probability of exceeding particular criteria in particular areas. The use of these three techniques also has the advantage of effectively requiring the data, both measurements and predictions, to be shown in context on a map. This avoids the alternate problems of ignoring information, and focussing attention on one or two particular measurements, or interpolating by eye without any mechanism to provide an independent check on the impression gained.

7.1.2   Speed and ease of use

The techniques discussed vary in complexity and none should be applied without an understanding of the approach used and the implementation steps required. However, some of the techniques can be both easy to use and produce distinct results shortly after sufficient data become available. For example, the Bayesian method will operate as a conventional Gaussian dispersion model until sufficient measurement results are available for an assimilation process to begin. Similarly ordinary kriging, although not useable until a few tens of measurements are available, is a comparatively simple sequence of procedures to apply. It also has the advantage that the additional work needed to implement the method requires the measurements and their spatial correlation to be critically examined before any kriging results are calculated. This examination, as well as ensuring the use of all the data, at least in the review stage, will potentially improve the quality of the data selected to support particular calculations even if no kriging then takes place.

Other indicators of the speed and ease of use of statistical techniques have been touched on, for example, when discussing the number of realisations required for a simulation to estimate contours indicating the probability of exceeding a threshold (see Sections 3.1.4.2, 6.4 and C3) or when discussing the use of kriging in conjunction with a process model (see Section 6.6). In the first case, like all statistical techniques, simulations improve as more data become available. An expected drawback of using simulations is the need to carry out a very large number of calculations. However, as indicated in Appendix C a reasonably stable result can be produced after only 100 realisations. In the second case, process models allow estimates of unmeasured quantities to be estimated not just at the sampling locations of the measured data used to derive the result but everywhere a kriged estimate can be produced, using a quick and simple procedure.

There is in general no computational bottleneck; the length of time taken by a calculation increases with the complexity of the calculation and the refinement of the answer required. The time taken for a calculation also conveniently matches to some

---

[*] It will improve on the estimates of simple models, as soon as a few tens of measurements are available unless the simple model is a near perfect description of the event eg the deposition arising from the accident at Tomsk.

degree the decision makers' needs and expectations, with initial estimates having higher uncertainty (due to lack of data) but being reliable enough to guide a robust early response. This is particularly true of the Bayesian assimilation process, as it does not require the initial exploratory data analysis of kriging techniques before it is applied and is no more difficult to use in practice than a standard Gaussian dispersion model that does not include a fitting procedure. However, in contrast, kriging with a trend was found difficult to use because of the large amount of exploratory data analysis and modelling required. It is therefore inappropriate to pursue the further development and application of this technique at this stage.

Very early decisions will, of necessity, be based on very few measurements and be dependent on estimates from simple process models, without the possibility of statistical support. As more data arrive, decisions will be required on which measurements to rely on, as the data as a whole are likely to be increasingly inconsistent with the predictions of a simple model. In these circumstances, there may be a gap before statistical methods are able to help (see Section 7.2.1). However, statistical methods are likely to be the best way of supporting and reconciling the assessment of a real world accident, allowing predictions that are more detailed with lower uncertainty to be made as the amount of monitoring data increases and refinements to countermeasures are considered.

The limitations on the application of the techniques considered in this study are lack of skills and data[*]. There should be no limitation introduced by the time taken to carry out a calculation once started. Given sufficient data and the skills to use it, the problem becomes primarily a question of organising the flow and manipulation of that data. The time consuming steps are making the data available in an appropriate form and gaining a general understanding of the event and the data so that the correct calculations with appropriate modelling assumptions are made. The use of databases and GIS will simplify the task, so that by the time there are sufficient data to influence estimates of what is occurring, beyond the scaling of process model predictions, calculations will be ready to proceed. To reach this stage would require some software development to streamline the process but off-the–shelf products that only need to be linked together and configured to handle the particular problems of environmental radioactivity measurements can meet most of these requirements. However, ensuring that there are sufficient skilled personnel to carry out the calculations is likely to be a greater obstacle to the effective use of these techniques. One approach to maintaining the appropriate skill base would to use the techniques more widely, for example in the assessment of contaminated land.

7.1.3 Use of real data

The techniques considered in SECTAR were specifically tested with past accident data. This has caused a number of problems due to the limited and generally messy nature of real data. A number of techniques, co-kriging being the prime example but also the use of process models and the techniques applied to the censored data from Tomsk, could

---

[*] Badly configured measurements will also affect the success of any calculation.

not be tested as effectively as was wished because the data or meta-data were
inadequate. The most dramatic failing was the inability to use the large number of
gamma dose rate measurements collected following the Windscale fire to support other
assessments (see Section 6.5). Windscale also highlighted the consequences of data
not being collected in a way that allowed for the most effective use of geostatistical
techniques (see Section 6.1). This having been said, the use of real data was necessary
to demonstrate the true worth of the methods tested when tackling realistic problems. It
also avoided the use of techniques with artificial data derived from a process model.
The techniques rely on modelling the spatial correlation between measurements and
are not likely to perform well where the data are excessively smooth (with the possible
exception of any arbitrary discontinuities added to make scenarios more realistic).
Where artificial data are used, simple dispersion models are likely to be a more
effective alternative.

7.1.4 Uncertainty

One of the most important attributes of the statistical methods under test is that they
provide information on the uncertainty of estimates and the likelihood of exceeding
thresholds. Simple geostatistical methods such as kriging provide a map of the kriging
variance that highlights regions where more data are required if estimates in those
areas are to be reliable. More substantially, it is possible using the simulation
techniques discussed in Sections 5.2 and 6.3 to provide probability contours that
indicate the likelihood of exceeding a particular value. For example, the probability of
exceeding the Windscale milk ban criterion and the related results of Figure 27, derived
using only 37 measurements and 100 realisations, demonstrate that simulation can
provide results that are both useful and timely. The technique can also be used to
provide contours of a chosen high percentile of each point estimate, to give decision-
makers a conservative view of the extent of contamination. For example, a 95%
confidence level could be selected instead of the mean estimate obtained from kriging.
Similar uncertainty estimates are available from a Bayesian calculation with the
possibility in this case of considering model parameter uncertainty as well as model
inadequacy. Although the approximate nature of the process model calculations used to
support emergency decision making is well understood the ability to apply some
confidence measures to predictions made under these conditions is a novel and
potentially very useful benefit of applying statistical methods.

7.1.5 Combining simple models and statistical techniques

The combined use of statistical techniques and simple process models has already
been discussed in the context of the Bayesian adapted Gaussian dispersion model
used in the SECTAR studies. However, it should be noted that the same basic
Bayesian formulation could also be used with other dispersion models or simple
process models. Similarly, geostatistical methods can also be used in conjunction with
simple modelling as demonstrated by the use of Windscale deposition data to derive
estimated concentrations in milk (see Section 6.6). The results were not particularly
good but, as discussed, this is not necessarily a limitation of the technique, and the
approach would have the merit of being both quick and easy to do (see Section 7.1.2).

7.2   Future developments

SECTAR has established that statistical techniques can be of use in the analysis of post-accident data in the short term. Several authors have used similar techniques to look at the distribution of deposition in the years following an accident as part of a radioecological analysis of the distribution and migration of radioactivity (Kanevsky *et al,* 1995; Dubois *et al*, 1998). However, there has been very little interest in using these techniques from the beginning of an event to improve accident response. This is partly a consequence of different communities having different perspectives and using different tools in their work. It is also a result of a misunderstanding of the requirements of rapid assessment shortly after a release ends, and the needs of the techniques employed. For example, geostatistical analyses are traditionally used to produce accurate results from large datasets. This is true in radioecology, where assessments are made without the overwhelming time constraints of accident response, and particularly in geostatistics' foundation discipline of geological prospecting. In this case, large financial commitments are considered by mining or oil companies partly based on such analyses but again without the acute time pressure of accident response. However, this does not preclude the successful use of geostatistical methods in situations where data are limited or when results must be obtained very quickly, although the accuracy of the results obtained under such circumstances will obviously be lower than in traditional uses of geostatistics. SECTAR has shown that the accuracy of the results is sufficient to enable a reasonable estimate of the extent of contamination to be made for the purposes of nuclear accident response.

7.2.1   Data

To critically test and evaluate statistical techniques requires a supply of data. The SECTAR project has been fortunate in having access to data from the accidents at Tomsk, Chernobyl (Bryansk) and Windscale. However, additional sources of data would be useful if new methods are to be tested and the work of SECTAR further developed. For example, work on temporal aspects as discussed in Section 7.2.3 will require time stamped data. The data within the Windscale database although often including time, usually as a date, is not sufficiently detailed to support such work. Data from accidental chemical releases and fires may help but the most likely source of the detailed information required would come from the enhanced routine monitoring of a pollutant source such as a chimney stack.

In addition to getting more measurements of the quantities to be assessed, getting more supporting data would also be of benefit. The SECTAR project has been unable to consider fully the use of supporting (secondary) data in the analysis of accident events. The analyses of both Tomsk and Windscale (see Sections 4.2 and 6.5) consider the use of supporting data. In the former case, this is provided by a model of atmospheric dispersion that captures the main features of the event and in the latter by a very limited supply of related radioactivity measurements. These examples are far from ideal and may not demonstrate the wider role secondary information could play. For example, the co-kriging trials conducted with Windscale data failed to demonstrate that the technique offered advantages that outweighed the burden of the procedure. However, the technique may have a use if sufficient supporting data could be made available eg a

large amount of dose rate information. Dose rates are likely to be available at an early stage and may be one way to help bridge the information gap and allow statistical techniques to be used at an early stage. However, the quantitative use of dose rate information to support other estimates would not be straightforward (see Section 7.2.3)[*].

An area where supporting information is likely to be important has recently been considered by NRPB (now HPA-RPD) and a report describing how radar rainfall information could be used to improve estimates of radioactive deposition has been published (Higgins and Jones, 2003). Unfortunately, a lack of data has prevented the techniques discussed being quantitatively tested. Although it may not be possible to test all aspects, if data from Chernobyl fallout measured in the UK were available, it might be possible for some trial calculations to be undertaken.

7.2.2 Uncertainty and process models

Simulation techniques have been used in the examination of Bryansk and Windscale but as discussed in Section 3.1.4.1 this technique has not been applied to highly trended data such as that of Tomsk. The potential of such techniques when the data are highly trended should be explored. However, the likely practical difficulty of developing geostatistical techniques that might work with limited amounts of highly trended data make further work in this area inappropriate.

When handling highly trended data Bayesian methods are more likely to be successful if a process model can adequately represent the trend. Although not explored within SECTAR, Bayesian techniques could also be used to provide probability contours and contours of high percentile estimates of the value of interest. If the process model were a good representation of the observed environmental behaviour, this would amount to an uncertainty analysis of some or all of the variables of the process model, conditioned by the available measurement results.

7.2.3 Temporal modelling

The use of statistical techniques within SECTAR has been largely confined to the assimilation of measurements made at different spatial locations. The data are taken to be representative of a single time or, in the case of Tomsk and Bryansk, at times sufficiently long after the deposition occurred for the time between measurements to be unimportant. Problems with using measurements made at different times were discussed in the context of the Windscale results in Section 6. In this case, data were corrected to a common time by using either simple radioactive decay or a process model to introduce an effective environmental half-life for the radioactive deposition. However, in many cases a more integrated approach would be useful, providing a more rapid and flexible assimilation procedure. There are schemes for including time as an extra dimension within a conventional geostatistical framework (De Cesare *et al*, 2002).

---

[*] The use of dose rates would present a number of theoretical and practical difficulties related to the changing radionuclide composition of the deposition in time and space. This would require additional information or modelling assumptions to be made.

However, a Bayesian[*] formulation may be a more appropriate perspective as it naturally evolves as more data become available. Whichever approach is chosen, it will extend the range of events that statistical techniques could help analyse, including an extended or long duration release.

One significant development that does not go as far as fully introducing time would be to develop the existing Bayesian tools produced for SECTAR to allow ground gamma dose rate information to be used. The problems with the Windscale gamma dose rate data available to SECTAR make the testing of a new formulation extremely difficult. However, it is known that gamma dose rates will be available from an early stage following a release and that they are quick and easy to take. They will therefore be one of the predominant measurement types likely to be available in the early stages of an accident, and, at least initially, will represent an important source of information about the release. Currently, theoretical limitations to the Bayesian formulation require the measurement data supporting a calculation either to precede or to be simultaneous with the quantities being predicted. Thus, time integrated air concentrations and deposition measurements can be used to support deposition estimates. However, the use of gamma dose rates to infer the level of deposition at other locations would require theoretical developments that enabled a posterior result (dose rate) to infer a prior result (deposition).

7.3   Recommendations

Section 7.2 has provided a list of investigations and technical developments that would provide both valuable underpinning of the statistical approaches advanced and useful additional capability. However, the priority must be to make the most promising of the techniques investigated under SECTAR readily available for use in an emergency. As discussed previously at a technical level this requires the linking together of software such as databases holding measurement and other information on the accident with a GIS and the necessary statistical tools. Unfortunately, although necessary, this is not a sufficient requirement to ensure that these tools will be available when required. To make the application of these techniques more likely in the early stages of a real event requires that staff maintain a familiarity with them. This requires the techniques to be in regular use either as part of a continuing research effort or as part of an emergency exercise programme. The former option is currently the most practical as exercises are generally of too short a duration for these techniques to come into their own. They would also be at a disadvantage in exercises, if results assumed for measurements were largely based on the predictions of a Gaussian dispersion model.

The priorities for future work are therefore to make the successful techniques of SECTAR (principally ordinary kriging and simulation in the far field[†], and Bayesian

---

[*] Or Bayesian kriging

[†] For a simple Gaussian dispersion model the near field can be interpreted as the distance the plume is likely to travel under constant atmospheric conditions. This may be as little as a few kilometres and Jones (1981) recommends that a simple Gaussian dispersion model (Clarke, 1979) should only used to estimate concentrations at distances less than 50 km from the source.

assimilation in the near field) easily useable by trained staff in the unlikely event of an accident and to continue with a research programme in this area. The latter will maintain the pool of trained staff and could be designed to improve the applicability of the techniques to limited duration exercises by further developing approaches that used air concentration and gamma dose rate measurements.

7.4 Conclusions

The SECTAR project has demonstrated the potential of statistical techniques to provide guidance in estimating the spread of radioactivity following an accident. It is recommended that a phased programme of further work be undertaken to capitalise on the results of SECTAR. The following simple steps are suggested:

a   Establish the ability to apply the technique of ordinary kriging rapidly to measurements of deposited radioactivity and the radioactivity then found in plants and animals after an accident as appropriate. This will fill the gap that currently exists in modelling at a local level contamination that has arrived from a distant source, complementing the use of long range dispersion models and aerial monitoring.

b   Extend the capability to the use of simulation in the far field once an ordinary kriging capability has been established. This work will complement the potential use of ensemble predictions by the meteorological office to predict the probability of contamination in particular areas.

c   Incorporate Bayesian assimilation as part of the standard repertoire of assessment tools. This will provide an opportunity for improved deposition estimates when either simple dispersion modelling or ordinary kriging may be unreliable eg in the former case beyond about 10 km from the site (Jones, 1981).

d   Prioritise further research activities in this area in line with the discussion of Section 7 in particular extending the capabilities of the Bayesian tools to use dose rate information.

Task a is, as discussed in Section 7.3, predominately a simple computing task that a reputable consultancy could undertake. Task b is slightly more complicated as common GIS software does not yet provide this functionality but guidance on the implementation and practical use of the utility routines used in this work could be provided to the implementing software consultancy by several research and university organisations that use these techniques. In a similar vein task c should be straightforward to implement with the assistance of the authors of the Bayesian method (Kennedy et al, 2002). However, some additional work may be required by them if the technique is to be used with alternative atmospheric dispersion models such as ADMS (CERC, 2002).

8 REFERENCES

Arnold L (1992). Windscale 1957. Macmillan.

Bourgeois C, Fourgoux M, Guilot L and Bergey C (1995). Aerial gamma mapping system for characterizing the impact of radioactive deposition on the environment. In: *Proceedings of an international symposium on environmental impact of radioactive releases*. IAEA, 117-122.

Brown J and Simmonds JR (1995). *FARMLAND A Dynamic Model for the Transfer of Radionuclides through Terrestrial Foodchains*. Chilton, NRPB-R273.

CEC (1989a). Council Regulation (Euratom) No 3954/87 laying down the maximum permitted levels of radioactive contamination of foodstuffs and feedingstuffs following a nuclear accident or any other case of radiological emergency. *Off J Eur Commun*, L371/11 (1987), amended by Council Regulation 2218/89. *Off J Eur Commun*, L211/1.

CEC. (1989b). Council Regulation (Euratom) No 944/89 laying down maximum permitted levels of radioactive contamination in minor foodstuffs following a nuclear accident or any other case of radiological emergency. *Off J Eur Commun*, L101/17.

CEC. (1990). Council Regulation (Euratom) No 770/90 laying down maximum permitted levels of radioactive contamination of feedingstuffs following a nuclear accident or any other case of a radiological emergency. *Off J Eur Commun*, L83/78.

CERC (2002). Atmospheric Dispersion Modelling System ADMS 3.1 Cambridge Environmental Research Consultants http://www.cerc.co.uk

Chamberlain AC (1959). Deposition of iodine-131 in Northern England in October 1957. *Quart J R Met Soc*, **85,** 350.

Chamberlain AC and Dunster HJ (1958). Deposition of Radioactivity in North-West England from the accident at Windscale. *Nature*, September 6 1958, 629-630.

Chamberlain AC (1981). Emission of Fission Products and Other Activities During the Accident to Windscale Pile No 1 in October 1957. AERE-M 3194, UKAEA Harwell.

Chiles J-P and Delfiner P (1999). Geostatistics Modeling Spatial Uncertainty. Wiley.

Clarke RH (1974). An analysis of the 1957 Windscale accident using the Weerie code. *Ann Nuclear Science and Engineering*, **1,** 73-82.

Clarke RH (1979). *A Model for Short and Medium Range Dispersion of Radionuclides Released into the Atmosphere*. Chilton, NRPB-R91.

Crick MJ and Linsley GS (1982). *An Assessment of the Radiological Impact of the Windscale Reactor Fire, October 1957*. Chilton, NRPB-R135.

Crick MJ and Linsley GS (1983). *An Assessment of the Radiological Impact of the Windscale Reactor Fire October 1957*. Chilton, NRPB-R135 Addendum.

Daniels WM and Higgins NA (2002). *Environmental Distributions and the Practical Utilisation of Detection Limited Environmental Measurement Data*. Chilton, NRPB-W13.

De Cesare L, Myers DE and Posa D (2002). FORTRAN programs for space-time modeling. *Computers & Geosciences*, **28,** 205-212.

Deutsch CV (1997). Direct Assessment of Local Accuracy and Precision. Geostatistics Wollongong '96 Volume 1 (edited by Baafi E Y and Schofield N A), 115 – 125. Kluwer Academic Publishers.

Deutsch CV (2000). Personal communication.

Deutsch CV and Journel AG (1998). GSLIB Geostatistical Software Library and User's Guide. Oxford University Press.

Dubois G, Malczewski J and De Cort M (1998). Spatial interpolation comparison 97. Special issue *J Geog Info and Dec Anal*, **2** (2).

Dunster HJ, Howells H and Templeton WL (1958). District Surveys following the Windscale Incident, October 1957. In Proceedings of the Second United Nations International Conference on the Peaceful Uses of Atomic Energy Vol 18 Waste Treatment and Environmental Aspects of Atomic Energy. United Nations, Geneva.

Goovaerts P (1998). Ordinary Cokriging Revisited. *Math Geol*, **30** (1), 21.

Higgins NA and Jones JA (2003). *Methods for interpreting monitoring data following an accident in wet conditions*. Chilton, NRPB-W34.

Isaaks EH and Srivastava RM (1989). An Introduction to Applied Geostatistics. Oxford University Press, Oxford.

Jones JA (1981). *The Estimation of Long range dispersion and deposition of continuous Releases of Radionuclides to Atmosphere*. Chilton, NRPB-R123.

Journel AG and Rossi ME (1989). When Do We Need a Trend Model in Kriging? *Math Geol*, **21** (7), 715-739.

Kanevsky M, Arutyunyan R, Bolshov L, Demyanov V, Savel'eva E and Haas T (1995). Report at 10th International Conference on Mathematical and Computer Modelling and Scientific Computing. 5-8 July 1995. Boston, MA, USA. Books of Abstracts pp 179.

Kennedy MC, O'Hagan A and Higgins N (2002). Bayesian analysis of computer code outputs. In Quantitative Methods for Current Environmental Issues (edited by Anderson CW, Barnett V, Chatwin PC and El-Shaarawi AH) 227-243, Springer-Verlag: London.

Krige DG (1951). A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand. Master's Thesis, University of Witwatersrand.

Matheron G (1962-63). Traite de geostatistique appliquee, Tome I; Tome II: Le krigeage. I:Memoires du Bureau de Recherches Geologiques et Minieres, No 14 (1962), Editions Technip, Paris; II: Memoires du Bureau de Recherches Geologiques et Minieres, No 24 (1963), Editions BRGM, Paris.

Ministry of Defence (1986). Defence Standard 05-58/Issue 2 Sampling Procedures and Tables for Inspection by Attributes of Isolated Lots.

Ministry of Defence (1985). Defence Standard 05-70/Issue 1 Guide To Defence Standard 05-58 Sampling Procedures and Tables for Inspection by Attributes of Isolated Lots.

Nisbet AF, Woodman RFM, Haylock RGE (1999). *Recommended Soil-to-Plant Transfer Factors for Radiocaesium and Radiostrontium for Use in Arable Systems*. Chilton, NRPB-R304.

Pannatier Y (1996). VARIOWIN: Software for Spatial Data Analysis in 2D. Springer-Verlag, New York.

Ryall D (2000). The Met Office dispersion Model, NAME NWP Gazette.

Simmonds JR (1985). *The Influence of season of the year on the transfer of radionuclides to terrestrial foods following an accidental release to atmosphere*. Chilton, NRPB-M121.

Shershakov VM, Vakulovski S M, Borodin RV, Vozzhennikov OL, Gaziev Ya L, Kosykh VS, Makhonto VS, Chumiciev VB, Korsakov AT, Martynenko VP and Godko A (1995). Analysis and prognosis of radiation exposure following the accident at the Siberian chemical combine Tomsk-7. *Rad Prot Dosim*, **59** (2), 93-126.

Xu W, Tran TT, Srivastava RM and Journel AG (1992). Integrating Seismic Data in Reservoir Modeling: the Collocated Cokriging Alternative. SPE 24742, 833-842. In Proceedings of 67th Annual Technical Conference of the Society of Petroleum Engineers.

Wackernagel H (1995). Multivariate Geostatistics. Springer-Verlag.

Yatsalo B (1997). Personal communication.

Appendix A   Ordinary kriging worked example

A1   INTRODUCTION

This example of an ordinary kriging estimate for a single location is intended to illustrate the steps in the calculation as performed by the GSLIB routine okb2d Deutsch and Journel, 1998). The data are selected from one of the random samples of Windscale iodine in milk data discussed in Section 6.1. All measurements are in units of μCi l^{-1}. Figure A1 shows the location at which an estimate is to be calculated and the values of the 9 measurements which fall within the isotropic search radius (11.5km) of this location. The variogram model parameters are those for the entire random sample, as used previously. They are: Range = 11.5km, sill = 0.0172, anisotropy angle (θ in Figure A2) = 125°, anisotropy ratio (a_{min}/a_{max}; see Figure A2) = 0.59.
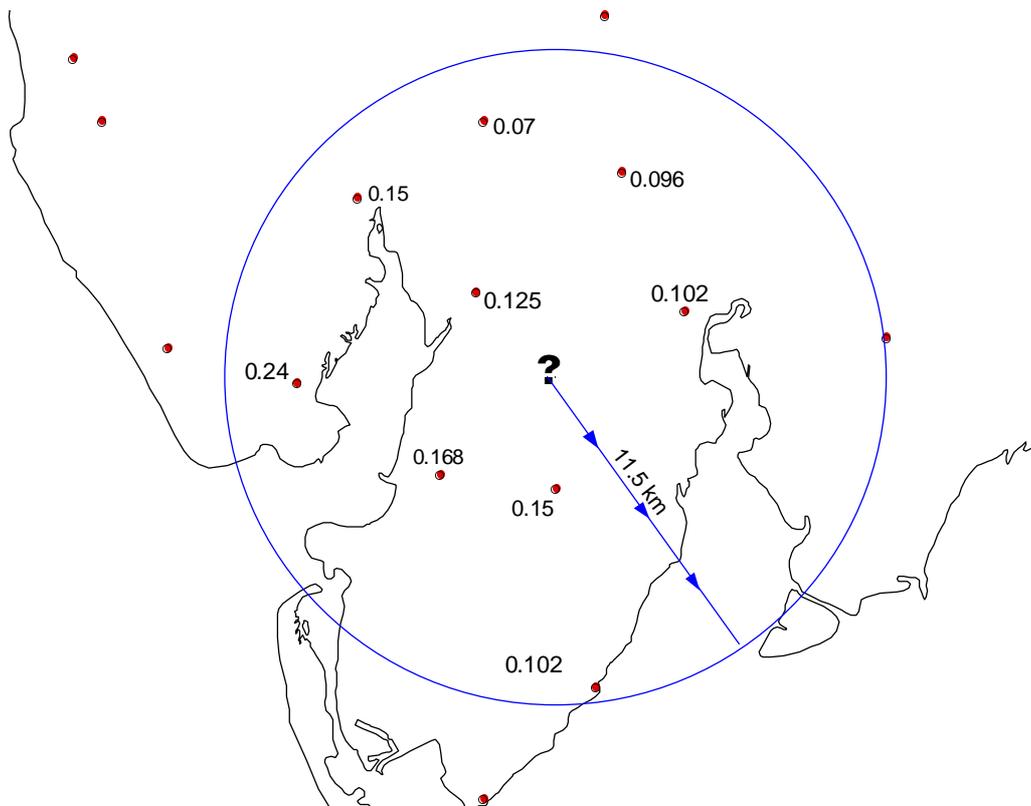

FIGURE A1 Example input data for ordinary kriging at a single location.

The ordinary kriging estimate for the location of interest is calculated from a weighted linear sum of the 9 measurement values, subject to the constraint that the weights sum to 1 to ensure that the estimate is unbiased. Since the measurement values are assumed to be realisations of a random function, the estimate formed in this way (Equation A1) is likewise a random function.

$$\hat{V}(x_0) = \sum_{i=1}^{9} w_i V(x_i) \qquad (A1)$$

$$R(x_0) = \hat{V}(x_0) - V(x_0) \qquad (A2)$$

$$Var\{R(x_0)\} = Var\{\sum_{i=1}^{9} w_i V(x_i) - V(x_0)\} \qquad (A3)$$

The unbiasedness of the estimate ensures that on average the estimation errors, defined by Equation A2, are zero. For a 'best' estimate, it is required that the *individual* errors are also as small as possible, which may be achieved by finding the set of weights which minimise the error variance, Equation A3. The variance is minimised by setting equal to zero its partial derivatives with respect to each of the weights, but with a constraint imposed by the unbiasedness condition. Thus, the method of Lagrange multipliers is used to solve the constrained minimisation problem.

The first step in the calculation is to obtain an expression for the error variance in terms of the covariances between pairs of measurement locations, C_{ij} and between the estimate location and each measurement location, C_{i0}. Assuming a stationary random function, that is, one whose mean is not location-dependent (at least over the search area) and whose covariance is only dependent on separation distance, Equation A4 may be derived for the error variance.

$$Var\{R(x_0)\} = Var\{V(x)\} + \sum_{i=1}^{9}\sum_{j=1}^{9} w_i w_j \tilde{C}_{ij} - 2\sum_{i=1}^{9} w_i \tilde{C}_{i0} \qquad (A4)$$

This equation uses covariances but it is more common in geostatistics to use the variogram to quantify the spatial autocorrelation. The covariance and variogram of a stationary random function are related by Equation A5. Given the variogram model it is therefore possible to calculate the covariances required in the above expression for the error variance.

$$C(h) = Var\{V(x)\} - \gamma(h) \qquad (A5)$$

The constrained minimisation of A4 uses partial differentiation with respect to the 9 weights and a Lagrange parameter to give, in this case, a set of 10 equations in 10 unknowns, which can be solved for the weights and the Lagrange parameter. The

solution is completed by substitution of the weights back into A1 to calculate the value of the random variable at the location of interest. The system of equations resulting from the minimisation can be compactly written in matrix form as shown in Equation A6.

$$
\begin{pmatrix}
\tilde{C}_{11} & \cdots & \tilde{C}_{19} & 1 \\
\vdots & \ddots & \vdots & \vdots \\
\tilde{C}_{91} & \cdots & \tilde{C}_{99} & 1 \\
1 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
w_1 \\
\vdots \\
w_9 \\
\mu
\end{pmatrix}
=
\begin{pmatrix}
\tilde{C}_{10} \\
\vdots \\
\tilde{C}_{90} \\
1
\end{pmatrix}
\tag{A6}
$$

That is, **C.w=D**. The matrix **C** is the matrix of covariances between pairs of measurement locations, and is therefore symmetric. The final row and column of this matrix are used to incorporate the constraint that the sum of weights must equal one. **D** is the matrix of covariances between the location at which an estimate is required and each of the 9 the measurement locations.

From Equation A6 the vector of weights may be obtained by a simple rearrangement, as shown in Equation A7.

w=C^{-1}D $\qquad$ (A7)

The calculation is complicated slightly if, as in this case, the phenomenon is anisotropic. The kriging equations assume an isotropic variogram model and so a transformation of the separation vectors **h**=(dx, dy) must be performed to work in the co-ordinate system defined by the axes of anisotropy. This transformation rotates the axes and stretches the scale along the anisotropy axes, as shown in Figure A2 and Equation A8.
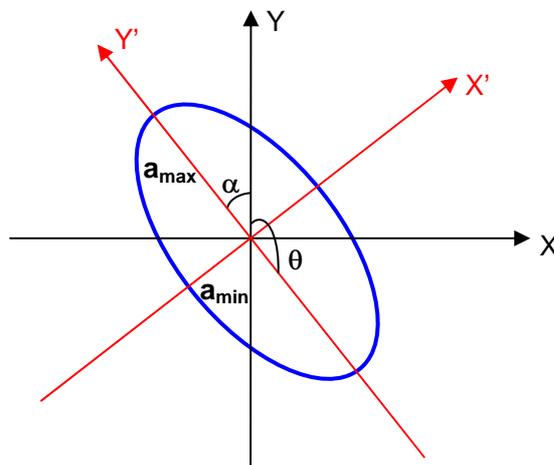

Figure A2 Illustration of anisotropy parameters.

$$\mathbf{h}' = \mathbf{T}\mathbf{R}\,\mathbf{h} \tag{A8}$$

Where

$$\mathbf{T} = \begin{pmatrix} \dfrac{1}{a_{min}} & 0 \\ 0 & \dfrac{1}{a_{max}} \end{pmatrix}$$

And

$$\mathbf{R} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$$

The isotropic (ie post-transformation by Equation A8) variogram model for the Windscale data in this example is given by Equation A9.

$$\gamma(h') = \begin{cases} 0.0172(1.5h' - 0.5h'^3) & h' \leq 1 \\ 1 & h' > 1 \end{cases} \tag{A9}$$

To calculate the covariance for any separation h' Equation A5 is used so that the variogram value given by Equation A9 must be subtracted from the variance of the random function, which is the variogram sill value 0.0172. The covariance matrix $\mathbf{C}$ and its inverse $\mathbf{C}^{-1}$ can now be calculated, as can the vector $\mathbf{D}$.

For the Windscale data in this example, these matrices and the resulting vector of kriging weights are given below:

$$\mathbf{C} = \begin{pmatrix}
0.0172 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0003 & 0.0005 & 0.0000 & 1.0000 \\
0.0000 & 0.0172 & 0.0013 & 0.0001 & 0.0062 & 0.0010 & 0.0002 & 0.0009 & 0.0001 & 1.0000 \\
0.0000 & 0.0013 & 0.0172 & 0.0000 & 0.0000 & 0.0005 & 0.0000 & 0.0000 & 0.0041 & 1.0000 \\
0.0000 & 0.0001 & 0.0000 & 0.0172 & 0.0013 & 0.0000 & 0.0050 & 0.0003 & 0.0000 & 1.0000 \\
0.0000 & 0.0062 & 0.0000 & 0.0013 & 0.0172 & 0.0018 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \\
0.0000 & 0.0010 & 0.0005 & 0.0000 & 0.0018 & 0.0172 & 0.0000 & 0.0000 & 0.0059 & 1.0000 \\
0.0003 & 0.0002 & 0.0000 & 0.0050 & 0.0000 & 0.0000 & 0.0172 & 0.0071 & 0.0000 & 1.0000 \\
0.0005 & 0.0009 & 0.0000 & 0.0003 & 0.0000 & 0.0000 & 0.0000 & 0.0172 & 0.0000 & 1.0000 \\
0.0000 & 0.0001 & 0.0041 & 0.0000 & 0.0000 & 0.0059 & 0.0000 & 0.0000 & 0.0172 & 1.0000 \\
1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.0000
\end{pmatrix}$$

$$\mathbf{C}^{-1} = \begin{pmatrix}
49.4983 & -5.4795 & -7.1869 & -7.0713 & -5.8535 & -6.0697 & -4.4469 & -8.2290 & -5.1616 & 0.1545 \\
-5.4795 & 63.9033 & -9.9174 & -3.0948 & -27.9068 & -5.5178 & -2.3242 & -7.9265 & -1.7363 & 0.0992 \\
-7.1869 & -9.9174 & 56.3261 & -6.0993 & -3.2922 & -1.0221 & -3.1855 & -5.2418 & -20.3798 & 0.0992 \\
-7.0713 & -3.0948 & -6.0993 & 58.8288 & -10.2294 & -4.5350 & -25.5159 & 2.2110 & -4.4941 & 0.1293 \\
-5.8535 & -27.9068 & -3.2933 & -10.2294 & 64.0594 & -10.6421 & -0.9514 & -3.7564 & -1.4265 & 0.1030 \\
-6.0697 & -5.5178 & -1.0221 & -4.5350 & -10.6421 & 62.9461 & -2.8779 & -4.5179 & -27.7638 & 0.1077 \\
-4.4469 & -2.3242 & -3.1855 & -25.5159 & -0.9514 & -2.8779 & 76.0031 & -34.5012 & -2.2000 & 0.0681 \\
-8.2290 & -7.9265 & -5.2418 & 2.2110 & -3.7564 & -4.5179 & -34.5012 & 66.0503 & -4.0885 & 0.1191 \\
-5.1616 & -1.7363 & -20.3798 & -4.4941 & -1.4265 & -27.7638 & -2.2000 & -4.0885 & 67.2506 & 0.09167 \\
0.1545 & 0.0992 & 0.1275 & 0.1293 & 0.1030 & 0.1077 & 0.0681 & 0.1191 & 0.0917 & -0.0027
\end{pmatrix}$$

Substituting into Equation A7 gives the weights, which are the first nine rows of the vector **w**. The tenth row in this vector is the Lagrange parameter, whose value is not needed when calculating the kriging estimate but is used if the kriging variance is also required. The kriging estimate is then obtained by substituting the weights into Equation A1, as shown in Equation A10, where the row vector **V** contains the measurement values within the search radius (see Figure 1) and the corresponding weights are in the column vector **w'**.

$$\mathbf{V.w'} = \begin{pmatrix} 0.102 & 0.125 & 0.102 & 0.240 & 0.150 & 0.070 & 0.168 & 0.150 & 0.096 \end{pmatrix} \begin{pmatrix} 0.0410 \\ 0.5275 \\ 0.1389 \\ 0.0590 \\ -0.1027 \\ 0.0256 \\ -0.0316 \\ 0.3355 \\ 0.0069 \end{pmatrix}$$

$$=0.130 \tag{A10}$$

That is, the ordinary kriging estimate for the location at the centre of the search area in Figure A1 is 0.130 μCi l^{-1}.

A2   REFERENCES

Deutsch CV and Journel AG (1998). GSLIB Geostatistical Software Library and User's Guide. Oxford University Press.

Appendix B Cokriging requirements

B1 INTRODUCTION

The discussion of Section 3.1.3 on cokriging is continued at greater depth in this Appendix. The intention is to highlight some of the particular difficulties associated with the technique.

B2 CHOICE OF NONBIAS CONDITIONS IN COKRIGING

The GSLIB routine (cokb3d) (Deutsch and Journel, 1998) used for cokriging in SECTAR offers three types of cokriging, which differ in their nonbias conditions. Simple cokriging is analogous to simple kriging in that there are no constraints on the weights applied to the primary or secondary variables. As with simple kriging, it is unlikely that a uniform mean can be assumed over the entire estimation area so this method is not very useful in practice, and ordinary cokriging is to be preferred. For p primary data at locations $\mathbf{x}_i$ and s secondary data at locations $\mathbf{x}_j$, traditional ordinary cokriging uses the nonbias conditions of Equation B1.

$$\sum_{i=1}^{p} w_i(\mathbf{x}_i) = 1$$

$$\sum_{j=1}^{s} w_j(\mathbf{x}_j) = 0 \tag{B1}$$

In this method, the local means are estimated and used by the kriging algorithm in the same way as for ordinary kriging with one variable. The disadvantage of constraining the sum of the secondary weights to equal zero is that some of them will necessarily be negative which increases the risk of obtaining negative estimates. The weights also tend to be small which results in the influence of the secondary variable being severely limited. The method also has the disadvantage that it cannot be used over local areas where there are no primary data because the constraint on the primary weights cannot be satisfied and the matrix of covariances will be singular. The third method is known as standardised ordinary cokriging, and it uses the nonbias conditions of Equation B2.

$$\sum_{i=1}^{p} w_i(\mathbf{x}_i) + \sum_{i=1}^{s} w_j(\mathbf{x}_j) = 1 \tag{B2}$$

The name of this method refers to the new rescaled ('standardised') secondary variables, which it uses. These have the same mean as the primary variable and are

calculated using Equation B3, where the v_j are the secondary data, and m_s and m_p are the means of the secondary and primary data respectively over the estimation area. These must be estimated by the user and supplied as input to the cokriging routine.

$$v_j^{rescaled}(\mathbf{x}_j) = v_j(\mathbf{x}_j) - m_s + m_p$$

(B3)

This rescaling clearly is only possible if primary and secondary data are either dimensionless or can be expressed in the same units.

The obvious difficulty with this method is the need to supply the primary and secondary means. Isaaks and Srivastava (1989) suggest that arithmetic averages of the primary and secondary sample data are suitable estimates for the means as long as the samples are not clustered. The local means (which will be the same for primary and rescaled secondary data) within each search neighbourhood will still be re-estimated by the standardised ordinary cokriging algorithm (Goovaerts, 1998). If data are clustered, it may be more appropriate to carry out cokriging separately for each area, using the appropriate sample mean for each cluster.

The advantages of using the single nonbias condition are twofold. Firstly, the likelihood of obtaining negative weights for secondary data (and consequently negative estimates which are often physically nonsensical) is reduced. Secondly, the magnitudes of the secondary weights are generally greater than they are in ordinary cokriging so the influence of the secondary data is increased. This is obviously beneficial when the primary data are scarce.

B3 VARIOGRAMS AND THE LINEAR MODEL OF COREGIONALISATION

If it has been possible to select appropriate data, the next issue is the creation and modelling of variograms to describe the spatial correlation and cross-correlation between the sample types. This process is considerably more involved than that described for a single variable in Section 3.1.1, both at the empirical variogram and modelling stages. For any of the three cokriging options in GSLIB (simple, ordinary or standardised; not collocated cokriging), empirical variograms for primary and secondary data are required. A cross-variogram which summarises the spatial cross-correlation between the two sample types is also needed, and for variables u, v at lag separation h this is defined by Equation B4.

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{ij} \approx h} (u_i - u_j).(v_i - v_j)$$

(B4)

This implies that a certain number of collocated primary and secondary data are needed whatever variety of cokriging is chosen. This situation may not always be easily

achieved for post-accident data as illustrated in Section 6.5, where 'artificially' collocated data are created for the cokriging trials undertaken.

Having created the three empirical variograms, they must be modelled subject to the constraint that the resulting cokriging matrix is positive definite (or equivalently, that the cokriging variance is always positive). The linear model of coregionalisation (Isaaks and Srivastava, 1989) is a method for doing this, and it is essentially a set of restrictions on the type and the parameters of the basic models which are used to fit the variograms[*]. Full mathematical descriptions are given in most geostatistics textbooks (eg Wackernagel, 1995), but the idea may be summarised as stating that the models for the individual primary and secondary variograms, and the cross-variogram, must be composed from a linear sum of the same basic variogram models. To ensure positive definiteness of the linear model, the nuggets and sills of these basic variogram models must satisfy the conditions of Equation B5.

$$\begin{vmatrix} c0_u & c0_{uv} \\ c0_{uv} & c0_v \end{vmatrix} > 0 \qquad \begin{vmatrix} c1_u & c1_{uv} \\ c1_{uv} & c1_v \end{vmatrix} > 0 \qquad \text{(B5)}$$

Here, $c0_u$ and $c0_v$ are the nuggets of the primary and secondary variograms and $c0_{uv} = c0_{vu}$ (it is symmetric) is the nugget of the cross-variogram. $C1_u$ and $c1_v$ are the sills of the primary and secondary variograms and $c1_{uv} = c1_{vu}$ is the sill of the cross variogram.

In general, if N structures were used to model the variogram (including the nugget), N determinants would need to be checked, one for the nugget plus one for the sills of each of the N-1 additional structures. It can be seen that these equations require that any structure included in the cross-variogram model must appear in the primary and secondary models, although the converse is not true; for example, including a nugget in the cross-variogram means that the primary and secondary variograms also need nuggets (which may be different), however, it is permissible for the primary and secondary variograms to have nuggets but not to include one in the cross-variogram model.

Modelling a coregionalisation can be time-consuming and there are a number of software packages available that automate the procedure to a greater or lesser extent. One such program, which was tested using the Windscale data, is Agromet (Bogaert *et al*, 1995), which will calculate appropriate sills and nuggets given the basic models and their ranges selected by the user. While this has the potential to save time and effort, the software was unable to work with anisotropy and so was unsuitable for the final cokriging experiments of SECTAR. The VarioWin software used for the single variable variograms was also used for the cokriging variograms. This effectively required the fitting to be done 'by eye', since the variograms and cross-variogram were modelled separately with the constraint that choosing nuggets and sills to satisfy the linear model of coregionalisation was more important than obtaining the 'best fit' for any individual

---

[*] This approach is sufficient but may not be necessary. Unfortunately, there is no generally available method of meeting the requirements with fewer imposed restrictions.

variogram. If cokriging is to be used in a practical accident situation, then variogram modelling will need to be automated to as great an extent as possible.

It was mentioned in Section 3.1.3 that collocated cokriging did not require the complicated variogram modelling procedure of the other varieties. Although lack of suitable data has prevented the practical investigation of collocated cokriging as part of SECTAR, the variogram requirements are outlined here for completeness. Collocated cokriging only needs a variogram for the primary data and a way to describe the cross-correlation. No secondary variogram is used because only the single collocated secondary datum is retained for each estimate location. An approximation for the cross-covariance can be calculated using a Markov model (Xu *et al*, 1992).

All investigations into cokriging under SECTAR were carried out using the Windscale dataset, as it was the only one to contain measurements of several different sample types and radionuclides. The data selection process and the experiments are described in Section 6.5.

B4   REFERENCES

Bogaert P, Mahau P, Beckers F (1995). The Spatial Interpolation of Agro-Climatic Data. Cokriging Software and Source Code User's Manual version 1.0b. Agrometeorology Series Working Paper, Number 12. FAO Rome, Italy. Environmental Information Management Service, Sustainable Development Department, Agrometeorology Group November 1995.

Deutsch CV and Journel AG (1998). GSLIB Geostatistical Software Library and User's Guide. OUP.

Goovaerts P (1998). Ordinary Cokriging Revisited. Math Geol, 30 (1), 21.

Isaaks EH and Srivastava RM (1989). An Introduction to Applied Geostatistics. Oxford University Press, Oxford.

Xu W, Tran TT, Srivastava RM and Journel AG (1992). Integrating Seismic Data in Reservoir Modeling: the Collocated Cokriging Alternative. SPE 24742, 833-842, in Proceedings of 67th Annual Technical Conference of the Society of Petroleum Engineers.

Wackernagel H (1995). Multivariate Geostatistics. Springer-Verlag

Appendix C The effect of simulation options

C1   INTRODUCTION

This section illustrates the effects of the choice of simulation parameters on the observed results.

C2   USE OF SIMPLE KRIGING

It was stated in Section 3.1.4.1 that theoretically, simple kriging should be used when carrying out simulations. However, it was also stated that ordinary kriging was thought a more appropriate choice for the Windscale data. Repeating the simulation experiments of Section 6.3 using simple kriging instead of ordinary kriging produced results, which were, for all practical purposes, identical.

C3   NUMBER OF REALISATIONS

An investigation was carried out into the effect of changing the number of realisations on the results obtained at specific locations in the Windscale analysis. Five points were selected at random from the simulation grid used in Section 6.3 and the statistics were calculated at regular intervals of 10 realisations up to 1000. Examples of how these statistics behaved as the number of realisations increased are shown in Figure C1. It can be seen that the statistics tended to settle down at the 300-realisations mark, this was generally the case for all points considered. The relative variance was the most erratic of the statistics, whereas the mean and the probability of exceeding the Windscale ban criterion converged to a limit relatively smoothly.

There appears to be very little to gain from using more than 300 realisations to produce statistics similar to those used in the Windscale simulations. However, for statistics on more extreme outliers, such as the 1st or 99th percentiles, more than 300 realisations should be considered.

To summarise, for an overall appreciation of the values and their uncertainties in the different regions of the map, it is reasonable to use statistics based on 100 realisations. For robust information on a smaller scale, 300 realisations would be sufficient, but more than this may be needed if highly detailed statistics are being calculated.

(a)

(b)
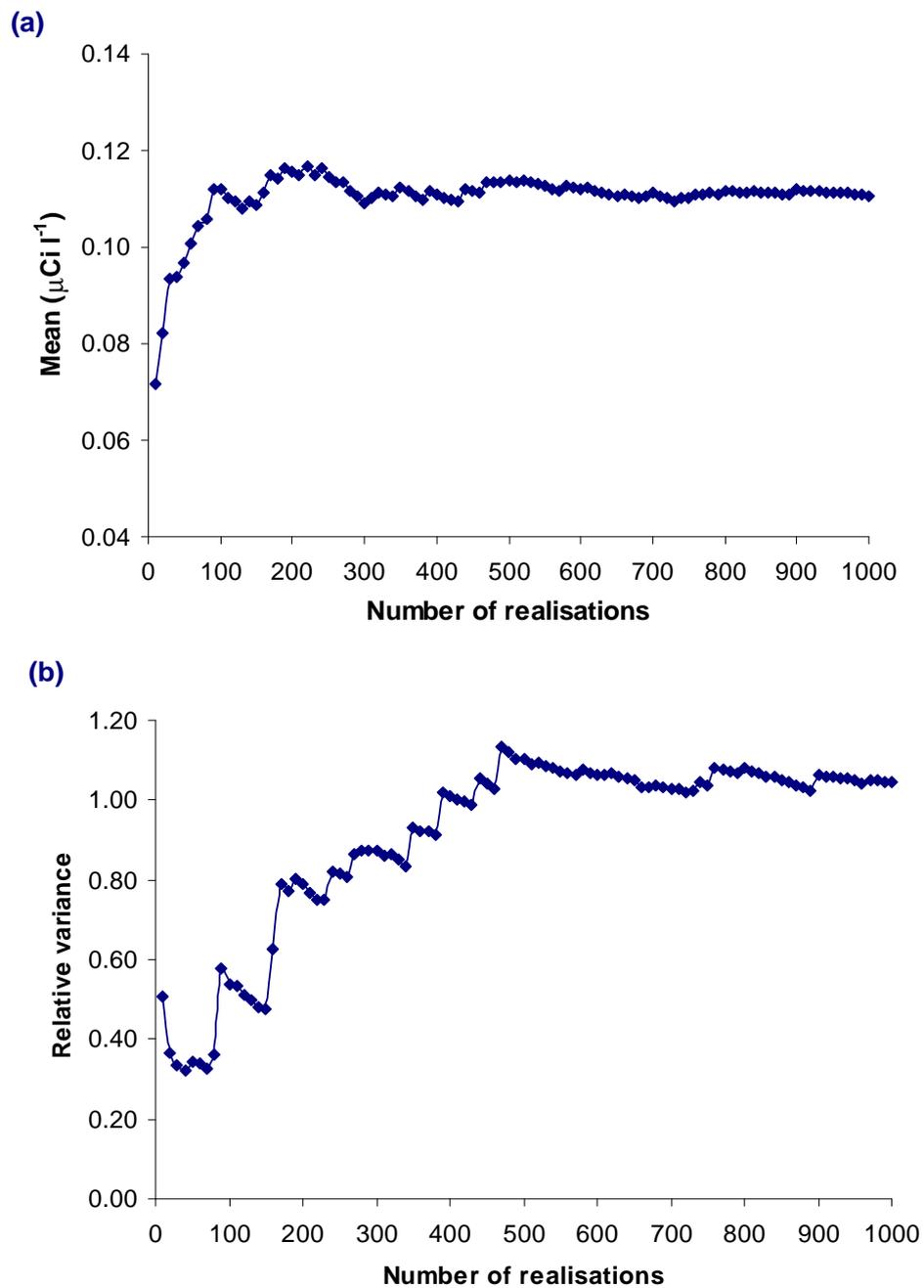
FIGURE C1 Effect of changing the number of simulation realisations on (a) the mean, and (b) the relative variance of the estimated milk activity concentration.
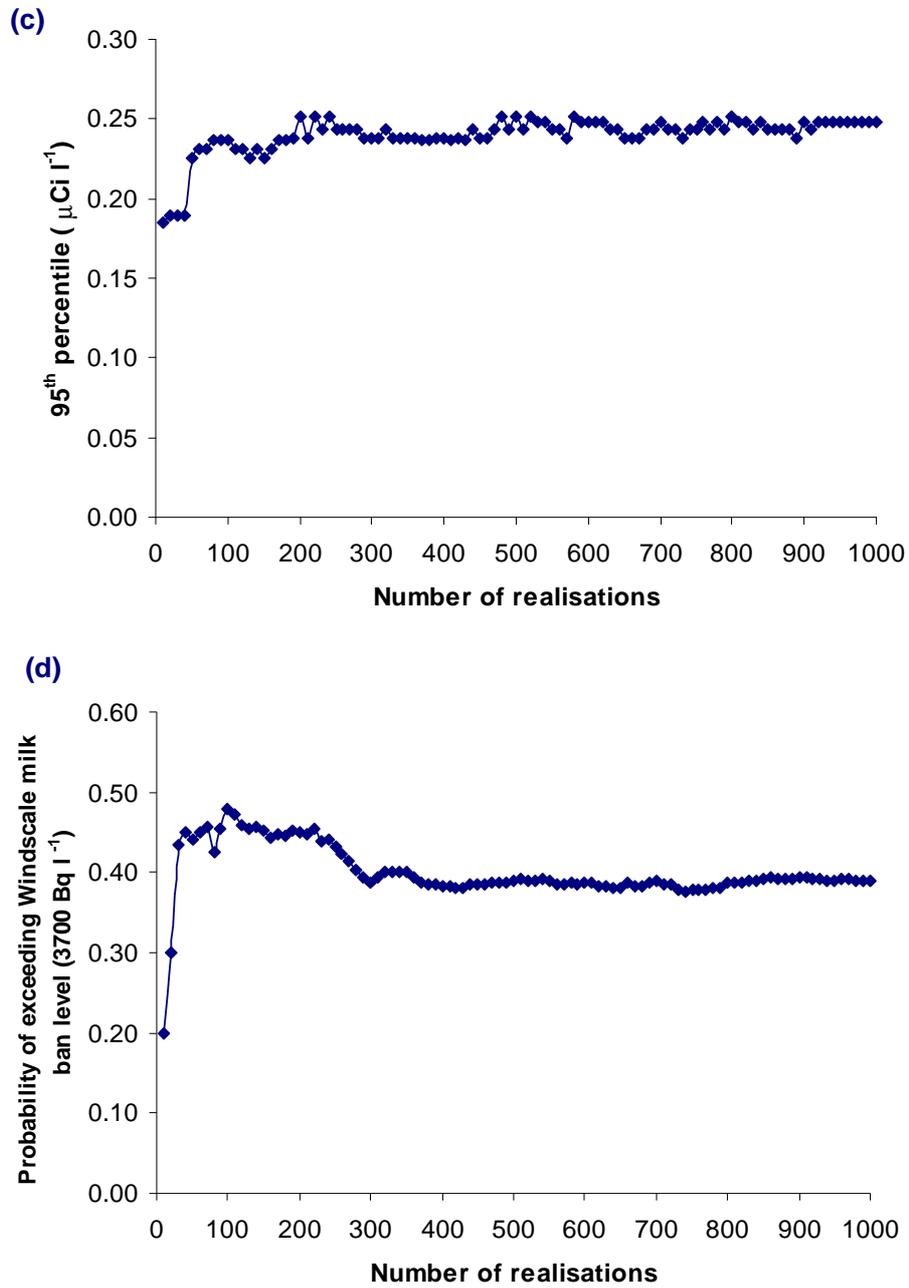
(c)

(d)

FIGURE C1 Effect of changing the number of realisations on (c) the 95th percentile of the estimated milk activity concentration, and (d) the probability of the estimates exceeding the Windscale milk ban criterion of 3700 Bq l^{-1} (0.1 μCi l^{-1}).

Appendix D   Windscale Database

D1   INTRODUCTION

The Windscale database was designed to hold data collected during the extensive programme of environmental surveys carried out after the 1957 fire at Windscale Pile No.1[*] for example, Chamberlain and Dunster (1958) and Dunster et al (1958). In the months following the accident, the results of the monitoring programme undertaken in the days and weeks following the fire were transferred to large punched cards, known as 'Paramount' cards[†]. Several copies of these cards were produced, and one set was subsequently acquired by NRPB (now HPA-RPD). The process of transferring these data to an ORACLE database has taken place over several years. The first version of the database organised the data in a structure that closely resembled the way it was stored on the punched cards. It was thought that this mirroring of the Paramount card structure would help to minimise the likelihood of errors being introduced on copying the data. However, it became clear that the data, once entered into the database, were awkward to retrieve. Thus, a major revision of the database structure was undertaken to make it easier to use. Most recently, data for meteorological measurements and polonium measurements, the latter of which were classified at the time of the accident (and hence not recorded on the Paramount cards) were added (Crabtree, 1959; Stewart et al, 1961). A comprehensive testing programme was also undertaken to establish the reliability of the record transfer for the various types of measurement.
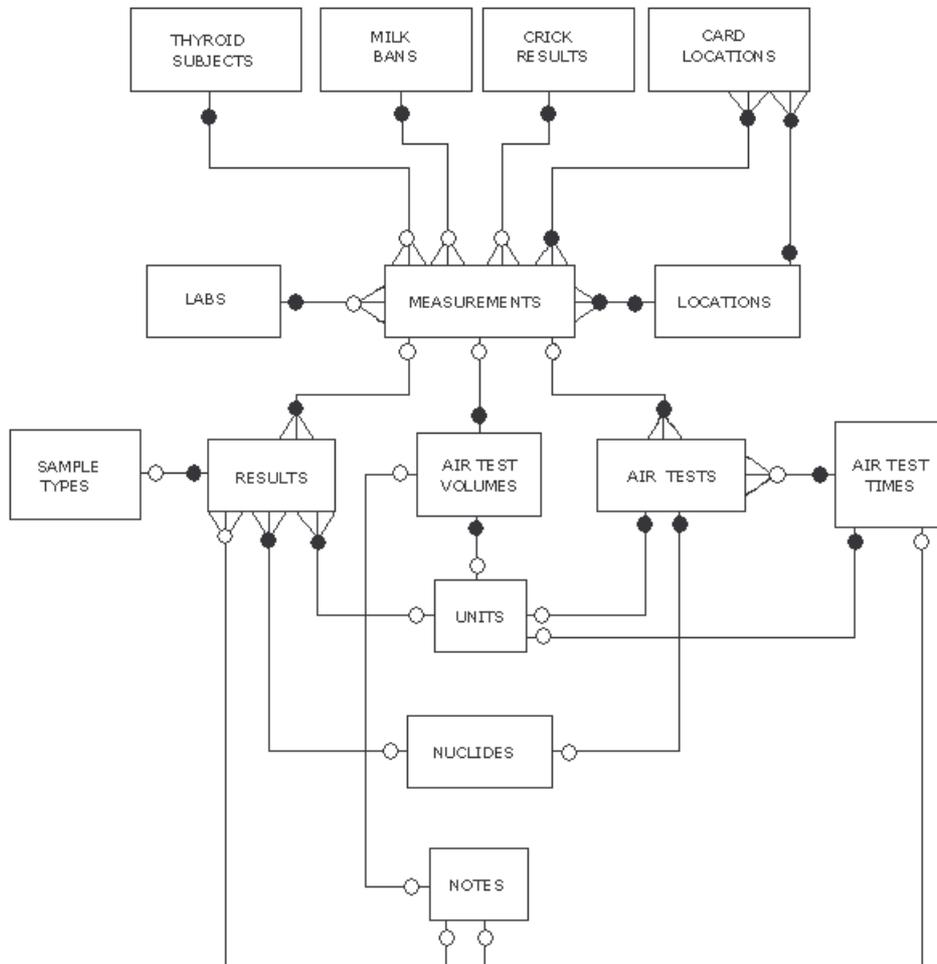
D2   STRUCTURE OF THE DATABASE

The Windscale database consists of 15 tables holding accident related data and 3 'QA' tables which are not directly connected to the main database. The QA tables hold details of any changes to the main database, the version number associated with the change and a record of any subsets of material from the database that has been supplied to others. The tables holding the accident data are arranged in the structure shown diagrammatically in Figure D1. The figure illustrates both the one-to-many relationships and the participation class of the tables. For example, the symbolic connection between the Results and Units tables illustrates that there may be many results with the same units in the Results table and that every row in the Results table

---

[*] A review of the history of the Windscale pile and accident can be found in Arnold (1992).
[†] Before the widespread availability of computers, the Paramount cards allowed data meeting particular criteria to be selected using a mechanical system of needles. The card holder and selector mechanism was not available at NRPB.

must have an associated unit. However, there may be units stored in the Units table, which are not used by the Results table.

FIGURE D1 Structure of database v2.3. The participation class of the relationship is denoted by the two circles. A black circle indicates mandatory participation by the adjacent table, ie, every entry in that table must refer to an entry in the table to which it is connected. An empty circle indicates optional participation, ie, the entries in the adjacent table are not necessarily referenced by the connected table.

A great deal of effort was devoted to the transfer of the information from the Paramount cards and other sources to the Oracle database and the subsequent testing of the accuracy of the transcription from the paper records. A further development which also proved successful was the connection of the database to a Geographic Information System (GIS). This allowed the location of the measurements to be checked for consistency. In the initial stages of database testing this assisted in the identification of locations that may have been wrongly reported or transcribed. As indicated in the introduction the results of sampling for radionuclides such as ^{210}Po, which were not

published at the time of the accident,[*] have also been included (Chamberlain, 1981; Crick and Linsley, 1982, 1983).

To help detect any systematic problems associated with transcribing specific types of measurements, the statistics quantifying the reliability of the transfer process were calculated for each sample type. These values were calculated as part of the testing procedure using a program based on Defence Standard 05-58/2 (Ministry of Defence, 1986). This program determined how many records should be tested and if errors were found how many more records had to be subsequently tested to achieve the set criteria.

D3 DATA QUALITY IN V2.3

A program supplied with the Ministry of Defence standard for testing isolated lots (Ministry of Defence, 1986) was used to determine the requirements for assessing of the reliability of the data transfer process that Version 2.3 of the Windscale database depended on. The program was run and the testing carried out for each sample class held in the database in case there were systematic errors that made particular sample classes more likely to contain errors than others. The program provides a number of test plans, which list the number of records to test each time and the number of allowed errors for each plan until the result is within the error bound selected.

The N records to test of the NLOT records of that type were selected using a random number generator. If the allowed number of errors for the particular test plan were exceeded, the lot was deemed to have failed the test. In this case, the tester must proceed to the second and subsequent test plans until the lot passed. However, before starting the next cycle of testing any data that had been found to be in error were corrected.

Given the number of plans used until the lot passed and the number of errors in the final sample, statistics can be compiled for the reliability of records in each sample class. The statistic quoted in Table D1 is the Lower Confidence Limit (LCL). This is the greatest expected percentage number of records in error for data of the given sample class. The test procedure statistics are formulated on the assumption that testing with replacement occurs in which case there is a 5% probability that the actual percentage error will be greater than that suggested by the LCL. However, testing without replacement was employed (which would result in more distinct records being tested when multiple lots had to be drawn) so that in these instances the LCL quoted would be an underestimate[†]. Additionally, the human thyroid, meat, miscellaneous and urine sample classes contain too few records for the program to generate a test plan. These were therefore tested exhaustively and their reliability can be stated as 100%, corresponding to an LCL of zero. The results of the testing are summarised in Table D1.

---

[*] These results were omitted from the Paramount cards.
[†] Multiple non-exhaustive testing did not occur in practice.

TABLE D1 Testing the reliability of the transcription of data from paper records to the Windscale Oracle database

Sample class	NLOT	Test number	Test conditions (sample to test, allowed no. of errors)	Errors[†]	Test result	LCL
Air	1948	1	(228,4)	2	PASS	2.772
Animal thyroid	117	1	(53,0)	1	FAIL	
Animal thyroid[*]	117	2	All remaining	0	PASS	0.000
Dairy	90	1	(46,0)	0	PASS	4.444
Gamma	2780	1	(235,4)	3	PASS	3.417
Human thyroid	4		ALL	0		0.000
Meat	7		ALL	0		0.000
MET	876	1	(180,3)	1	PASS	2.511
Milk	3310	1	(238,4)	3	PASS	3.414
Miscellaneous	11		ALL	2		0.000
Urine	40		ALL	1		0.000
Vegetable	1124	1	(186,3)	2	PASS	3.292
Water	150	1	(78,1)	1	PASS	4.000

[†]All errors found were corrected

[*]No advantage in limiting the number tested

D4   DATA SOURCES

Table D2 below lists the origin of all records in version 2.3 of the Windscale database. The polonium measurements are taken from Stewart et al (1961), which was declassified in 1962. Paramount cards with locations, measurement dates and sample type (air) information exist for these measurements, however no actual results are recorded on the cards as they were typed in 1958 when the polonium data was still classified. This report also contains measurement data for mainland Europe but these have not been included, as the values recorded are very low in comparison with the UK measurements and unlikely to be of any practical use in data analysis.

Table D2 Sources of Windscale data

Measurement number	Source	Notes
1-669	File 1, Paramount cards	
670-730	File 6, Paramount cards	Special location cards
731-1774	File 1, Paramount cards	
1775-3141	File2, Paramount cards	
3142		NOT IN DATABASE
3143-4158	File 3, Paramount cards	
4159-5278	File 4, Paramount cards	
5279-7243	File 5, Paramount cards	
7244-7396	Files 6, Paramount cards	Special location cards
7400-7434	Stewart et al (1961)	Table 1 data in pCi m^{-3} and held in air_tests table. Derived data from Crick and Linsley (1983), in Bq s m^{-3}, held separately in Crick_results table.
7435-7440	Crick and Linsley (1983)	Table 4
7441-7443		NOT IN DATABASE
7444	A C Chamberlain (1981), File 4 Corney	^{90}Sr, ^{210}Po in milk, Corney, 16/10/57: 147, 26 dpm/l
7445-7459	D.V Booker (1958)	Table 1
7460-7493	D.V Booker (1958)	Table 4
7494-7530	D.V Booker (1958)	Table 5
7531-7537	Stewart et al (1961)	Table 1
8001-8900	Document E18 in file B1055/02	Met data for the Windscale area on 10th – 11th October, 1957

D5   REFERENCES

Arnold L (1992). Windscale 1957 – Anatomy of a nuclear accident, Macmillan.

Booker DV (1958). Physical measurements of activity in samples from Windscale. (AERE HP/R 2607).

Chamberlain AC (1981). Emission of Fission Products and Other Activities During the Accident to Windscale Pile No 1 in October 1957. AERA M 3194.

Chamberlain AC and Dunster HJ (1958). Deposition of radioactivity in North-West England from the accident at Windscale. *Nature*, **182**, 629-630.

Crick MJ and Linsley GS (1982). *An Assessment of the Radiological Impact of the Windscale Reactor Fire, October 1957*. Chilton, NRPB R-135.

Crick MJ and Linsley GS (1983). Addendum to R-135.

Crabtree J (1959). The travel and diffusion of the radioactive material emitted during the Windscale accident. *Quart J Roy Met Soc*, **85**, 363-371.

Dunster HJ, Howells H and Templeton WL (1958). District surveys following the Windscale accident. 2nd Conference on the Peaceful Uses of Atomic Energy, 18, 296.

Ministry of Defence (1986). Sampling procedures and tables for inspection by attributes of isolated lots 05-58/issue 2.

Stewart NG, Crooks RN, Fisher EMR (1961). Measurements of the Radioactivity of the cloud from the accident at Windscale: Data submitted to the IGY. (AERE-M 857).